Progress in Probabilistic Logic Programming

Luc De Raedt
(using many slides of Angelika Kimmig)

DTAI
MACHINE LEARNING

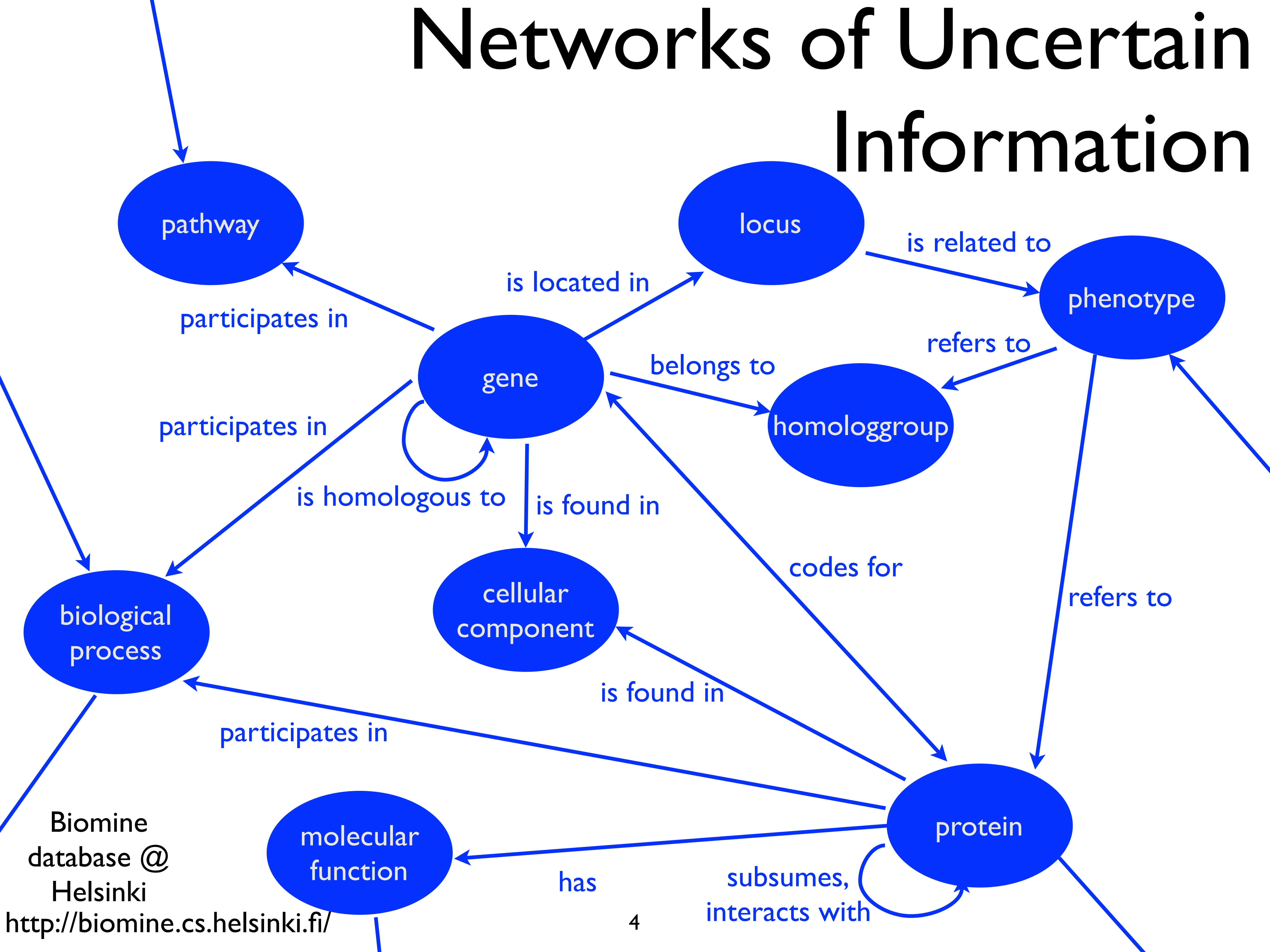
KU LEUVEN

Overview

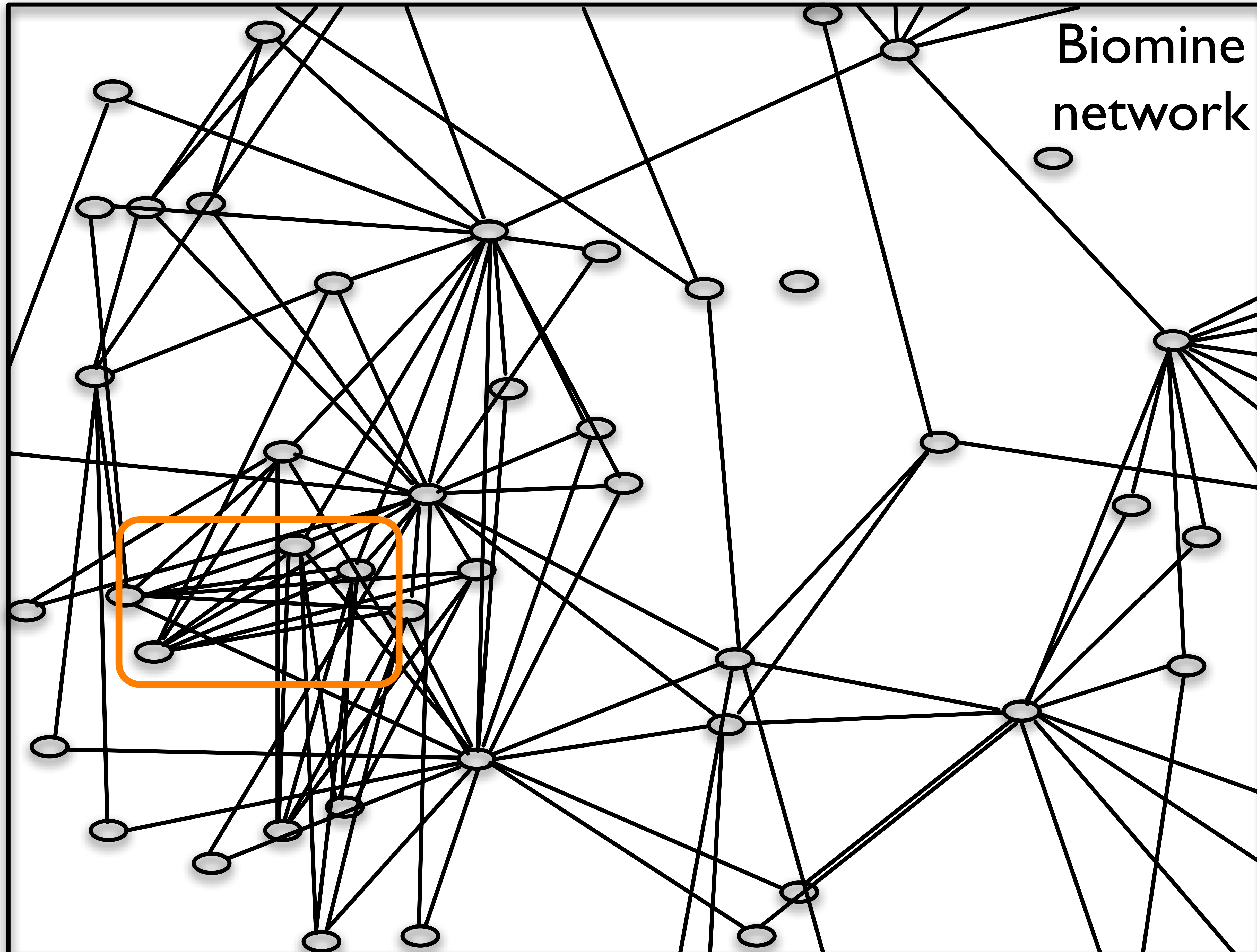
- Part I : Recap basic PLP framework & relation to alternative frameworks
- Part II : Inference (Basics)
- Recent issues
  - Part III : decision theoretic Problog
  - Part IV: Dynamics & Continuous distributions for Relational Tracking (in Robotics)
  - Part V : Probabilistic rule learning (ProbFOIL)

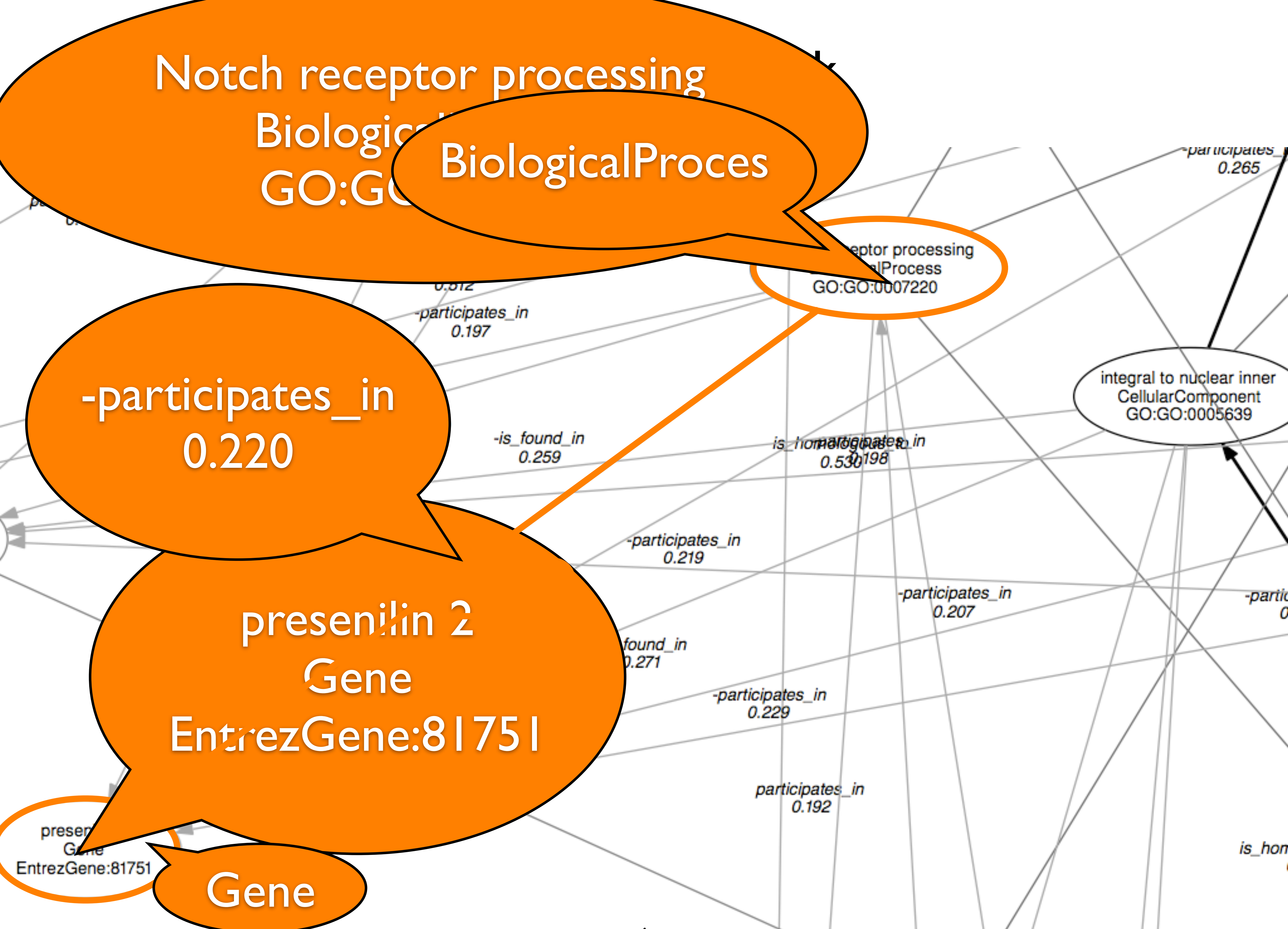
PART 1: Intro to PLP

Networks of Uncertain Information

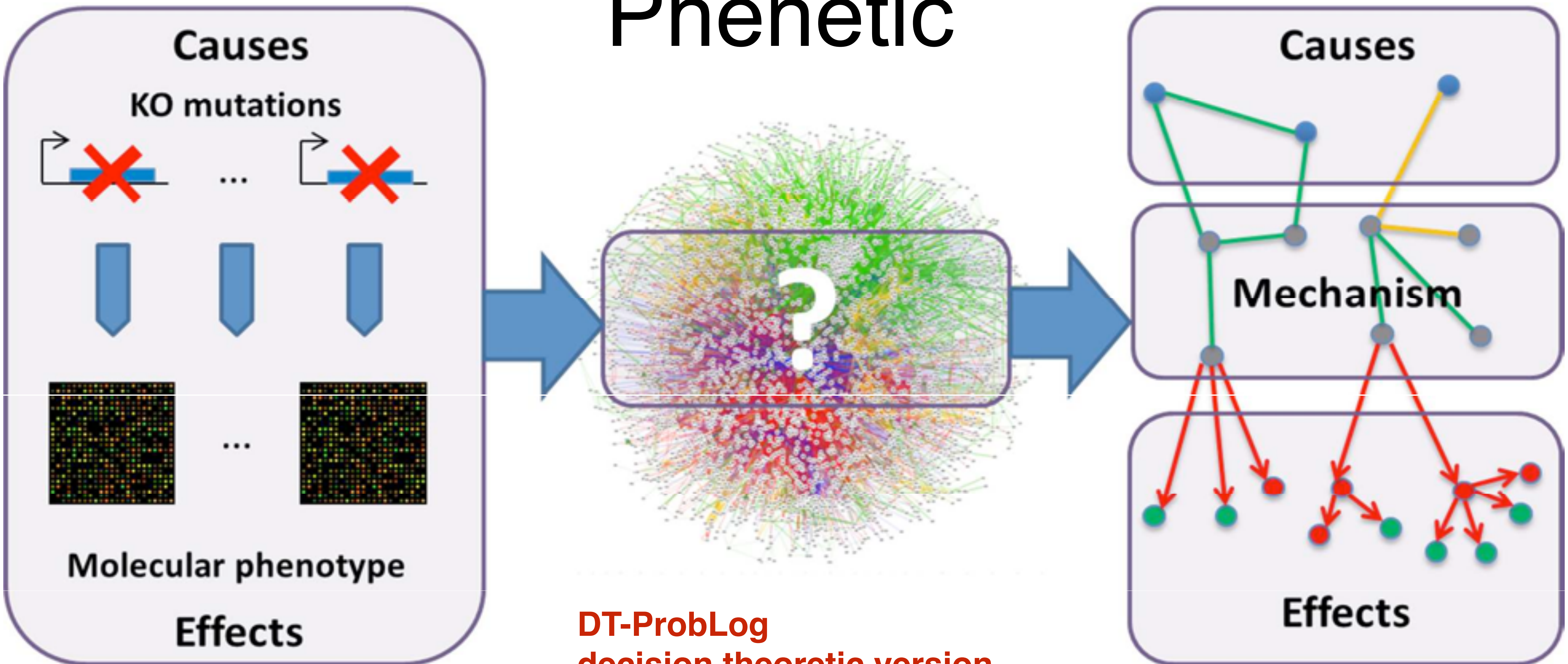


Biomine database @ Helsinki
http://biomine.cs.helsinki.fi/


Biomine network


Notch receptor processing
BiologicalProces
-participates_in
0.220
presenilin 2
Gene
EntrezGene:81751
Gene
GO:GO:0007220
integral to nuclear inner
CellularComponent
GO:GO:0005639
0.265
-participates_in
0.197
-is_found_in
0.259
-participates_in
0.219
-participates_in
0.207
-participates_in
0.229
participates_in
0.192
EntrezGene:81751

Phenetic



DT-ProbLog
decision theoretic version

- Causes: Mutations
  - All related to similar phenotype
- Effects: Differentially expressed genes
- 27 000 cause effect pairs

- Interaction network:
  - 3063 nodes
    - Genes
    - Proteins
  - 16794 edges
    - Molecular interactions
    - Uncertain

- Goal: connect causes to effects through common subnetwork
  - = Find mechanism
- Techniques:
  - DTProbLog
  - Approximate inference

Can we find the mechanism connecting causes to effects?

[De Maeyer et al., Molecular Biosystems 13, NAR 2015]

Graphs & Randomness

ProbLog, Phenetic, Prism, ICL, Probabilistic Databases, ...

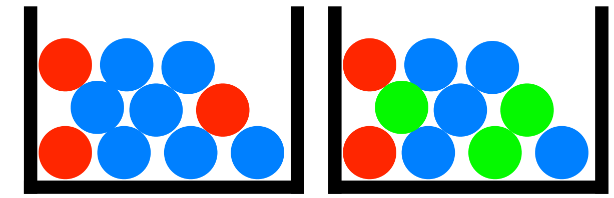
- all based on a “random graph” model

Stochastic Logic Programs, ProPPR, PCFGs, ...

- based on a “random walk” model
- connected to PageRank
- not the subject of this talk !

ProbLog by example:

A bit of gambling



- toss (biased) coin & draw ball from each urn
- win if (heads and a red ball) or (two balls of same color)

probabilistic fact: heads is true with probability 0.4 (and false with 0.6)

annotated disjunction: first ball is red with probability 0.3 and blue with 0.7

probabilistic choices

```
0.4 :: heads.

0.3 :: col(1,red); 0.7 :: col(1,blue) <- true.
0.2 :: col(2,red); 0.3 :: col(2,green);
                   0.5 :: col(2,blue) <- true.
```

annotated disjunction: second ball is red with probability 0.2, green with 0.3, and blue with 0.5

```
win :- heads, col(_,red).
win :- col(1,C), col(2,C).
```

logical rule encoding background knowledge

consequences

Questions

```
0.4 :: heads.

0.3 :: col(1,red); 0.7 :: col(1,blue) <- true.
0.2 :: col(2,red); 0.3 :: col(2,green); 0.5 :: col(2,blue) <- true.

win :- heads, col(_,red).
win :- col(1,C), col(2,C).
```

marginal probability

- Probability of **`win`**?

conditional probability

- Probability of **`win`** given **`col(2,green)`**?

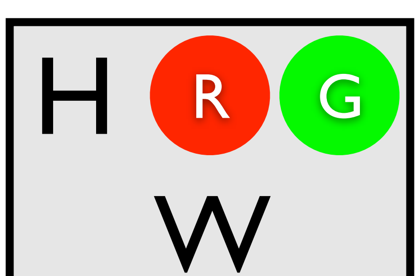
- Most probable world where **`win`** is true?

MPE inference

Possible Worlds

```
0.4 :: heads.

0.3 :: col(1,red); 0.7 :: col(1,blue) <- true.
0.2 :: col(2,red); 0.3 :: col(2,green); 0.5 :: col(2,blue) <- true.

win :- heads, col(_,red).
win :- col(1,C), col(2,C).
```

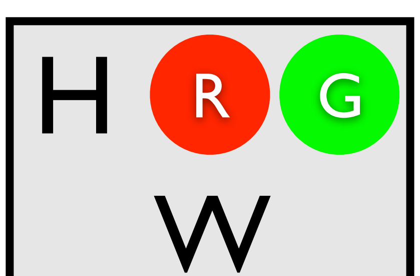
$0.4 \times 0.3 \times 0.3$



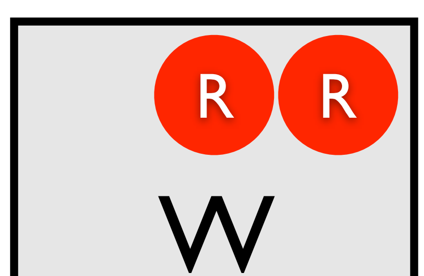
Possible Worlds

```
0.4 :: heads.

0.3 :: col(1,red); 0.7 :: col(1,blue) <- true.
0.2 :: col(2,red); 0.3 :: col(2,green); 0.5 :: col(2,blue) <- true.

win :- heads, col(_,red).
win :- col(1,C), col(2,C).
```

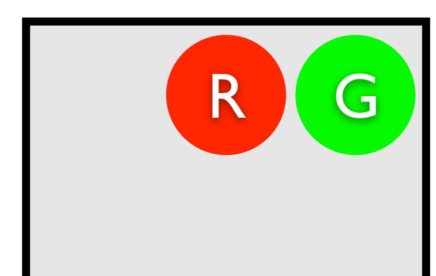
$0.4 \times 0.3 \times 0.3$



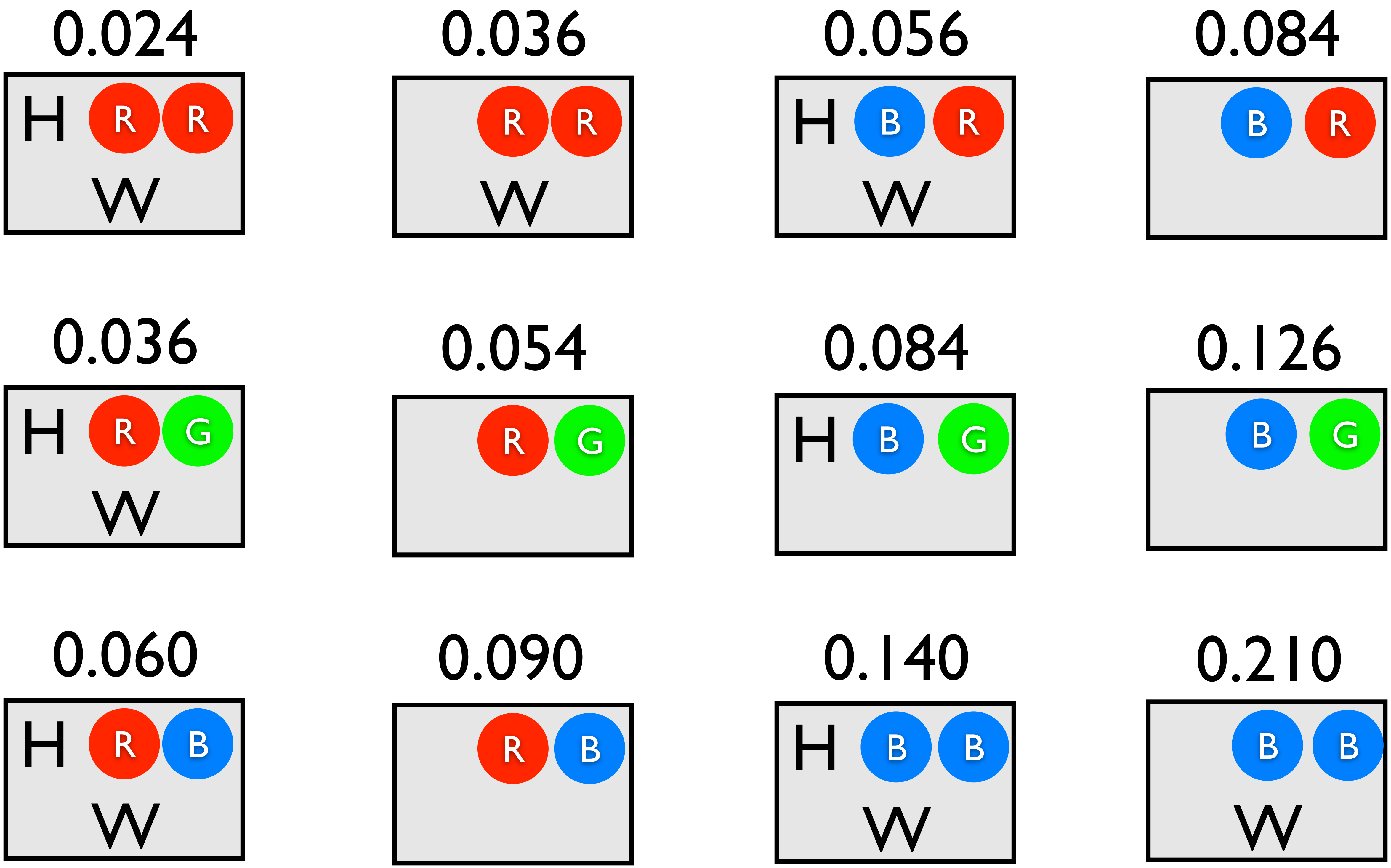
$(1-0.4)\times 0.3 \times 0.2$



$(1-0.4)\times 0.3 \times 0.3$



All Possible Worlds



Distribution Semantics

(with probabilistic facts)

[Sato, ICLP 95]
[Poole, AIJ 92]

query

sum over possible worlds where Q is true

$$P(Q) = \sum_{F \cup R \models Q} \underbrace{\prod_{f \in F} p(f) \prod_{f \notin F} 1 - p(f)}_{\text{probability of possible world}}$$

subset of probabilistic facts

Prolog rules

cProbLog: constraints on possible worlds

```
weight(skis,6).
weight(boots,4).
weight(helmet,3).
weight(gloves,2).

P::pack(Item) :-
   weight(Item,Weight),
   P is 1.0/Weight.

excess(Limit) :- ...

not excess(10).
pack(helmet) v pack(boots).
```

distribution over all possible worlds

normalized distribution over restricted set of possible worlds

sbhg e(10)	sb g e(10)	sbh e(10)	sb
s hg e(10)	s g	s h	s
bhg	b g	bh	b
hg	g	h	

constraints
as FOL formulas
treat as evidence

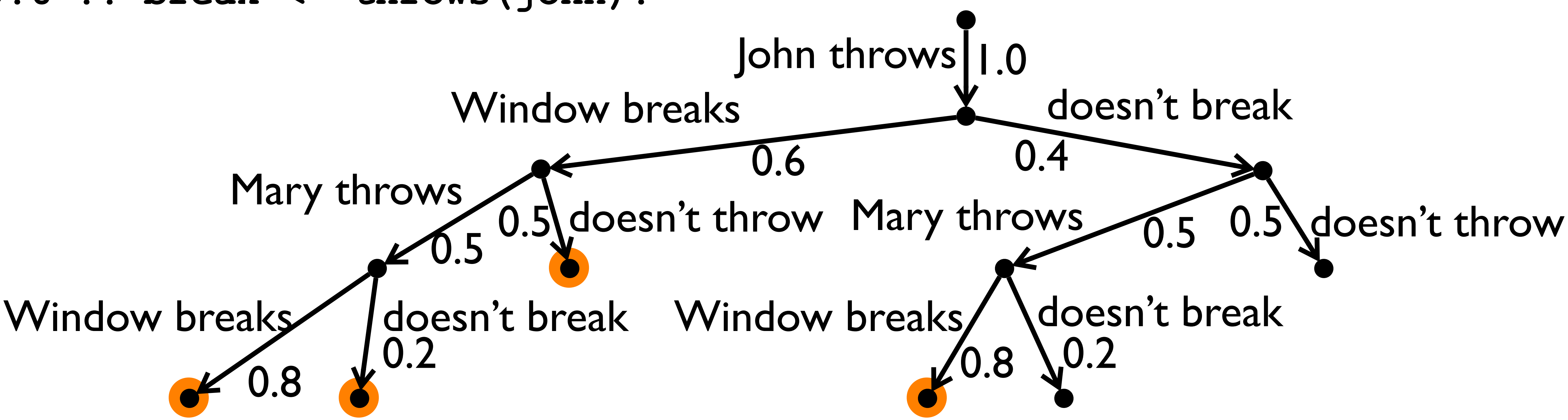
[Fierens et al, PP 12; Shterionov et al]

[Vennekens et al, ICLP 04]

Alternative view: CP-Logic

```
throws(john).
0.5::throws(mary).

0.8 :: break <- throws(mary).
0.6 :: break <- throws(john).
```

probabilistic causal laws



P(break)=0.6×0.5×0.8+0.6×0.5×0.2+0.6×0.5+0.4×0.5×0.8

CP-logic [Vennekens et al.]

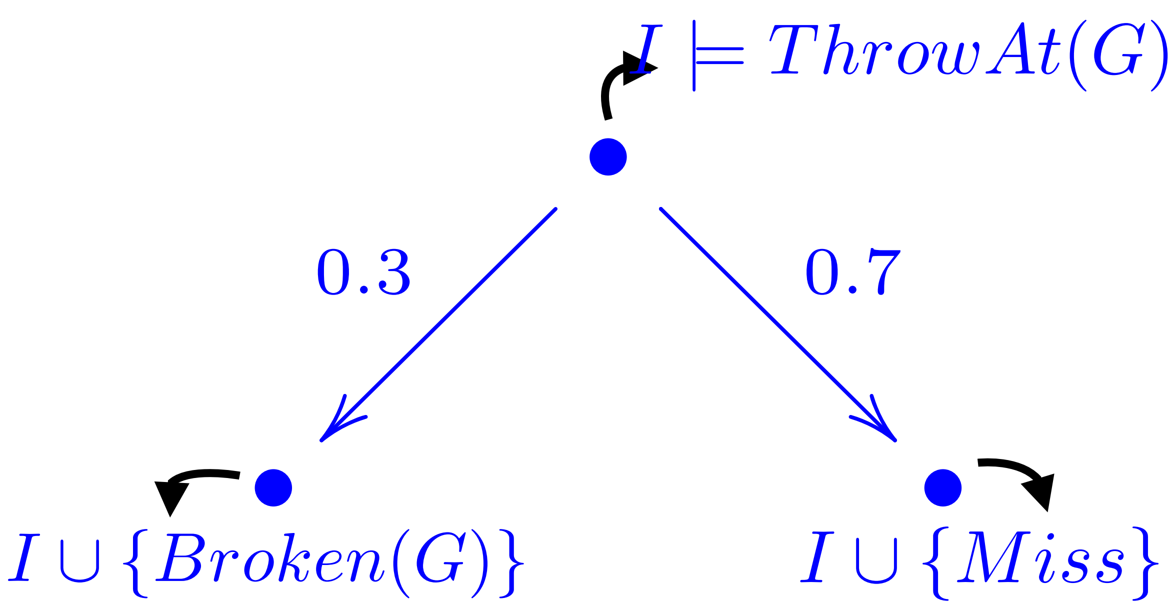
E.g., "throwing a rock at a glass breaks it with probability 0.3 and misses it with probability 0.7"

$$(Broken(G){:}0.3) \vee (Miss\ 0.7) \leftarrow ThrowAt(G).$$

ote that the actual non-deterministic event ("rock flying at glass") is implicit

Slides CP-logic courtesy Joost Vennekens

Semantics



(Broken(G) 0.3) ∨ (Miss 0.7)
← ThrowAt(G)

Probability tree is an execution model of theory iff:

- Each tree-transition matches causal law
- The tree cannot be extended
- Each execution model defines the same probability distribution over final states

Slides CP-logic courtesy Joost Vennekens

Distributional Clauses (DC)

- Discrete- and continuous-valued random variables

LDR 27 Jul 2014, 10:51
Defines a generative process (as for CP-logic)
Tree can become infinitely wide
Sampling

random variable ... distribution

```
length(Obj) ~ gaussian(6 ... (Obj,glass).
stackable(OBot,OTop) :- ...
        ≃length(OBot) ≥ ≃length(OTop),
        ≃width(OBot) ≥ ≃width(OTop).
ontype(Obj,plate) ~ finite([0 : glass, 0.0024 : cup,
                            0 : pitcher, 0.8676 : plate,
                            0.0284 : bowl, 0 : serving,
                            0.1016 : none])
                 :- obj(Obj), on(Obj,O2), type(O2,plate).
```

...**mparing** values of random variables

random variable with discrete distribution

[Gutmann et al, TPLP 11; Nitti et al, IROS 13]

Probabilistic Logic Programming

Distribution Semantics [Sato, ICLP 95; Poole]:
probabilistic choices + logic program
→ distribution over possible worlds

OVERVIEW paper [Kimmig, De Raedt, Arxiv, MLJ in press]

e.g., PRISM, ICL, ProbLog, LPADs, CP-logic, ...

multi-valued switches

probabilistic alternatives

probabilistic facts

annotated disjunctions

causal-probabilistic laws

Problog

- extends probabilistic databases
- is a probabilistic programming language
- is a statistical relational learning / AI system

Probabilistic databases

programming versus database query language
different types of queries

ProducesProduct

Company	Product	P
sony	walkman	0.96
microsoft	mac_os_x	0.96
ibm	personal_computer	0.96
microsoft	mac_os	0.9
adobe	adobe_indesign	0.9
adobe	adobe_dreamweaver	0.87
...	...	...

HeadquarteredIn

Company	City	P
microsoft	redmond	1.00
ibm	san_jose	0.99
emirates_airlines	dubai	0.93
honda	torrance	0.93
horizon	seattle	0.93
egyptair	cairo	0.93
adobe	san_jose	0.93
...	...	...

```
select x.Product, x.Company
from ProducesProduct x, HeadquarteredIn y
where x.Company=y.Company and
y.City='san_jose'
```


[Example from Suciu et al 2011]

Probabilistic Programs

- Distributional clauses / PLP similar in spirit
  - to e.g. BLOG, ... but embedded in existing logic and programming language
  - to e.g. Church but use of logic instead of functional programming ...
    - natural possible world semantics and link with prob. databases.
    - somewhat harder to do meta-programming

Markov Logic

Key differences

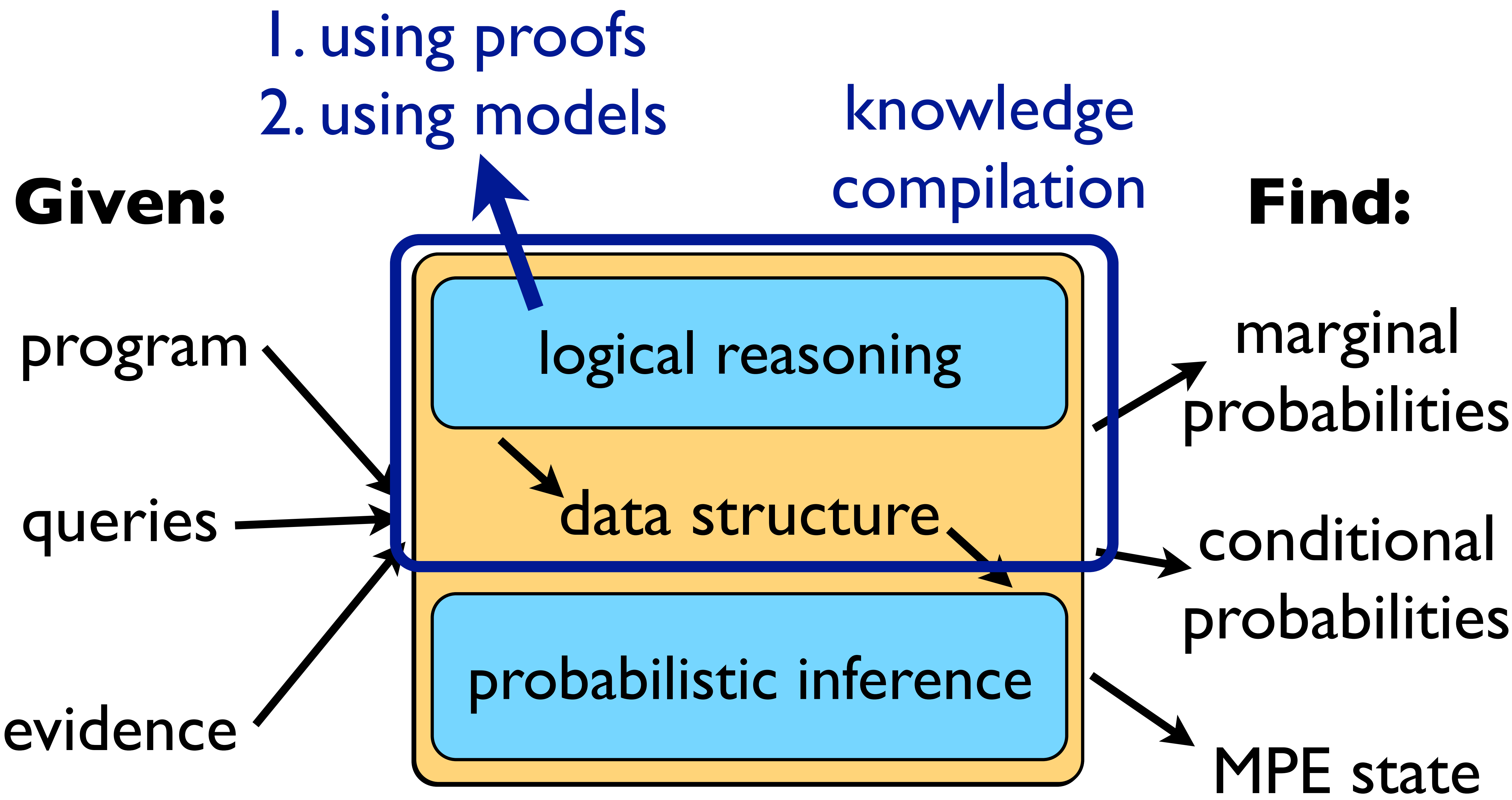
- programming language
- Pro(b)log uses least-fix point semantics
  - can express transitive closure of relation
  - this cannot be expressed in FOL (and Markov Logic), requires second order logic
  - p(X,Y) :- p(X,Z), p(Z,Y).
  - p(X,Y) :- edge(X,Y).   edge(1,2).

PART II: Inference

Inference in PLP

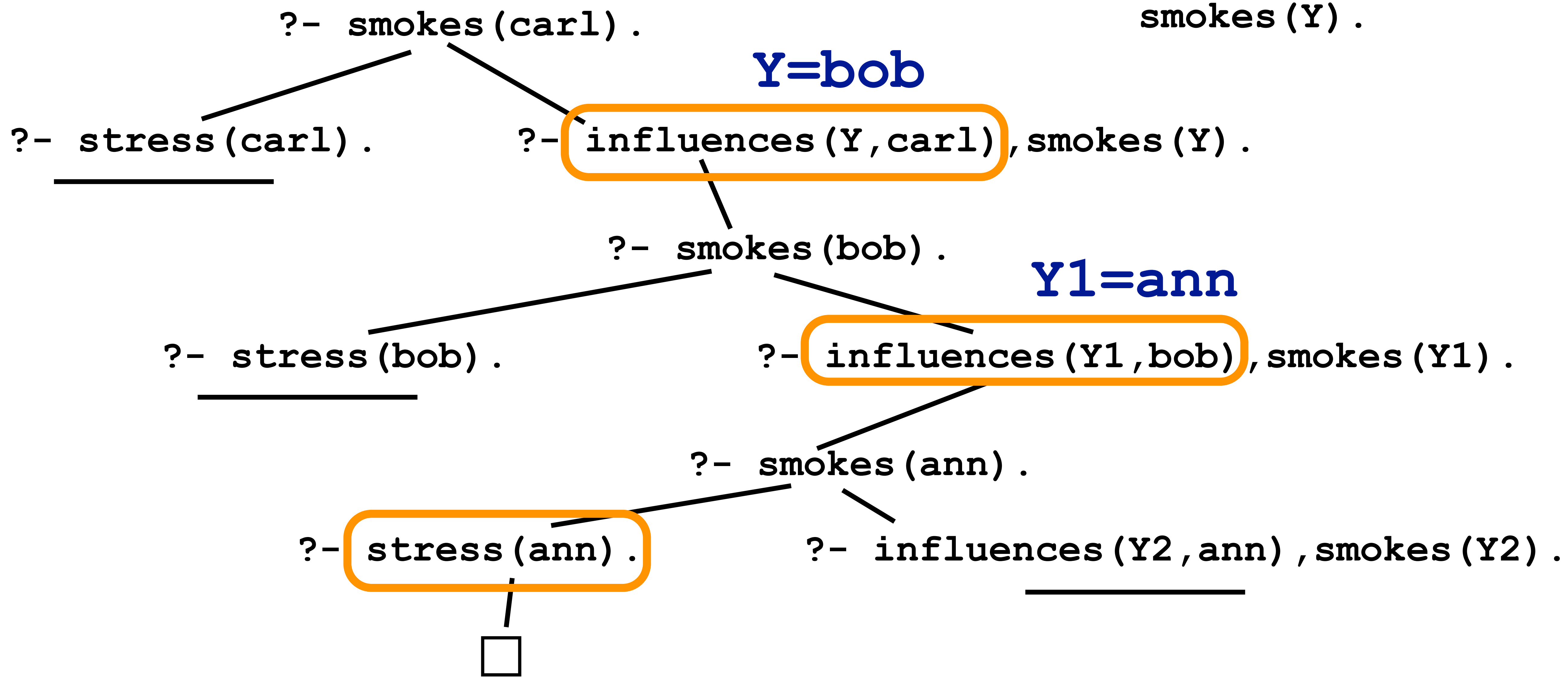
- As in Prolog and logic programming
  - proof-based
- As in Answer Set Programming
  - model based
- As in Probabilistic Programming
  - sampling

Answering Questions



Logical Reasoning: Proofs in Prolog

```
stress(ann).
influences(ann,bob).
influences(bob,carl).

smokes(X) :- stress(X).
smokes(X) :-
      influences(Y,X),
      smokes(Y).
```



proof = facts used in successful derivation:

`influences(bob,carl)&influences(ann,bob)&stress(ann)`

Proofs in ProbLog

```
0.8::stress(ann).
0.6::influences(ann,bob).
0.2::influences(bob,carl).

smokes(X) :- stress(X).
smokes(X) :-
      influences(Y,X),
      smokes(Y).
```

```
?- smokes(carl).
?- stress(carl).
?- influences(Y,carl),smokes(Y).
Y=bob
?- smokes(bob).
?- stress(bob).
?- influences(Y1,bob),smokes(Y1).
Y1=ann
?- smokes(ann).
?- stress(ann).
?- influences(Y2,ann),smokes(Y2).
□
```

influences(bob,carl)&influences(ann,bob)&stress(ann)

probability of proof = $0.2 \times 0.6 \times 0.8 = 0.096$

Proofs in ProbLog

```
0.8::stress(ann).
0.4::stress(bob).
0.6::influences(ann,bob).
0.2::influences(bob,carl).

smokes(X) :- stress(X).
smokes(X) :-
    influences(Y,X),
    smokes(Y).
```

```
?- smokes(carl).
?- stress(carl).
?- influences(Y,carl),smokes(Y).
```

Y=bob

```
?- smokes(bob).
```

```
influences(bob,carl)
  & stress(bob)
```

```
?- stress(bob).
```

0.2×0.4
$= 0.08$

$\square$

Y1=ann

```
?- influences(Y1,bob),smokes(Y1).
?- smokes(ann).
?- stress(ann).
?- influences(Y2,ann),smokes(Y2).
```

$\square$

```
influences(bob,carl)
& influences(ann,bob)
  & stress(ann)
```

$0.2 \times 0.6 \times 0.8$
$= 0.096$

proofs overlap!
cannot sum probabilities
(disjoint-sum-problem)

Disjoint-Sum-Problem

possible worlds

solution: knowledge compilation

```
infl(bob,carl) &   infl(ann,bob) &   st(ann) & \+st(bob)
infl(bob,carl) &   infl(ann,bob) &   st(ann) &   st(bob)
infl(bob,carl) & \+infl(ann,bob) &   st(ann) &   st(bob)
infl(bob,carl) &   infl(ann,bob) & \+st(ann) &   st(bob)
infl(bob,carl) & \+infl(ann,bob) & \+st(ann) &   st(bob)
...
```

0.0576
0.0384
0.0256
0.0096
0.0064

`influences(bob,carl) & stress(bob)`

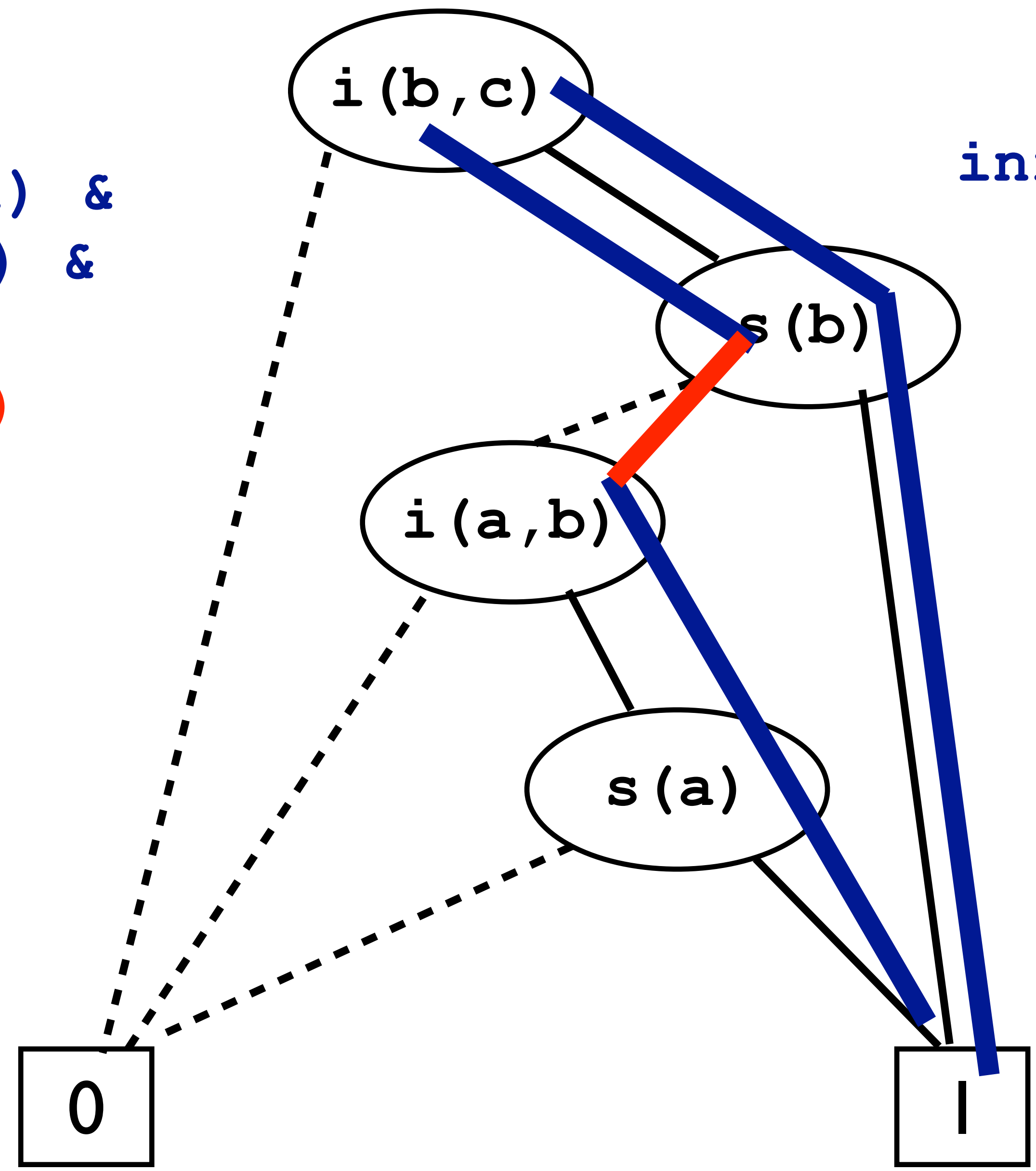
$\sum = 0.1376$

sum of proof probabilities: 0.096+0.08 = 0.1760

Binary Decision Diagrams

[Bryant 86]

influences(bob,carl) & influences(ann,bob) & stress(ann) & not stress(bob)



influences(bob,carl) & stress(bob)

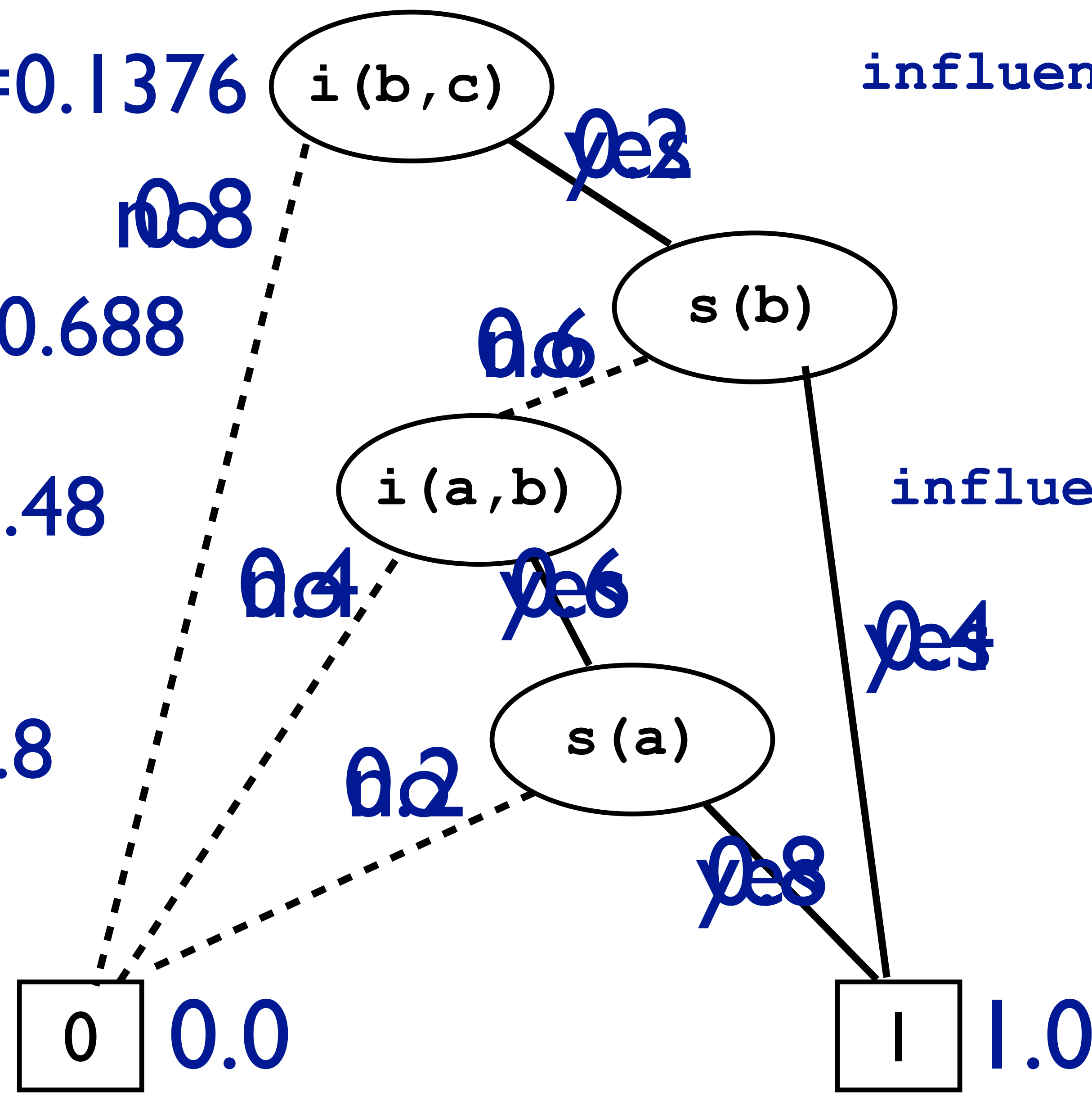
Binary Decision Diagrams

0.8×0.0+0.2×0.688=0.1376

0.6×0.48+0.4×1.0=0.688

0.4×0.0+0.6×0.8=0.48

0.2×0.0+0.8×1.0=0.8



`influences(bob,carl)?`

`stress(bob)?`

`influences(ann,bob)?`

`stress(ann)?`

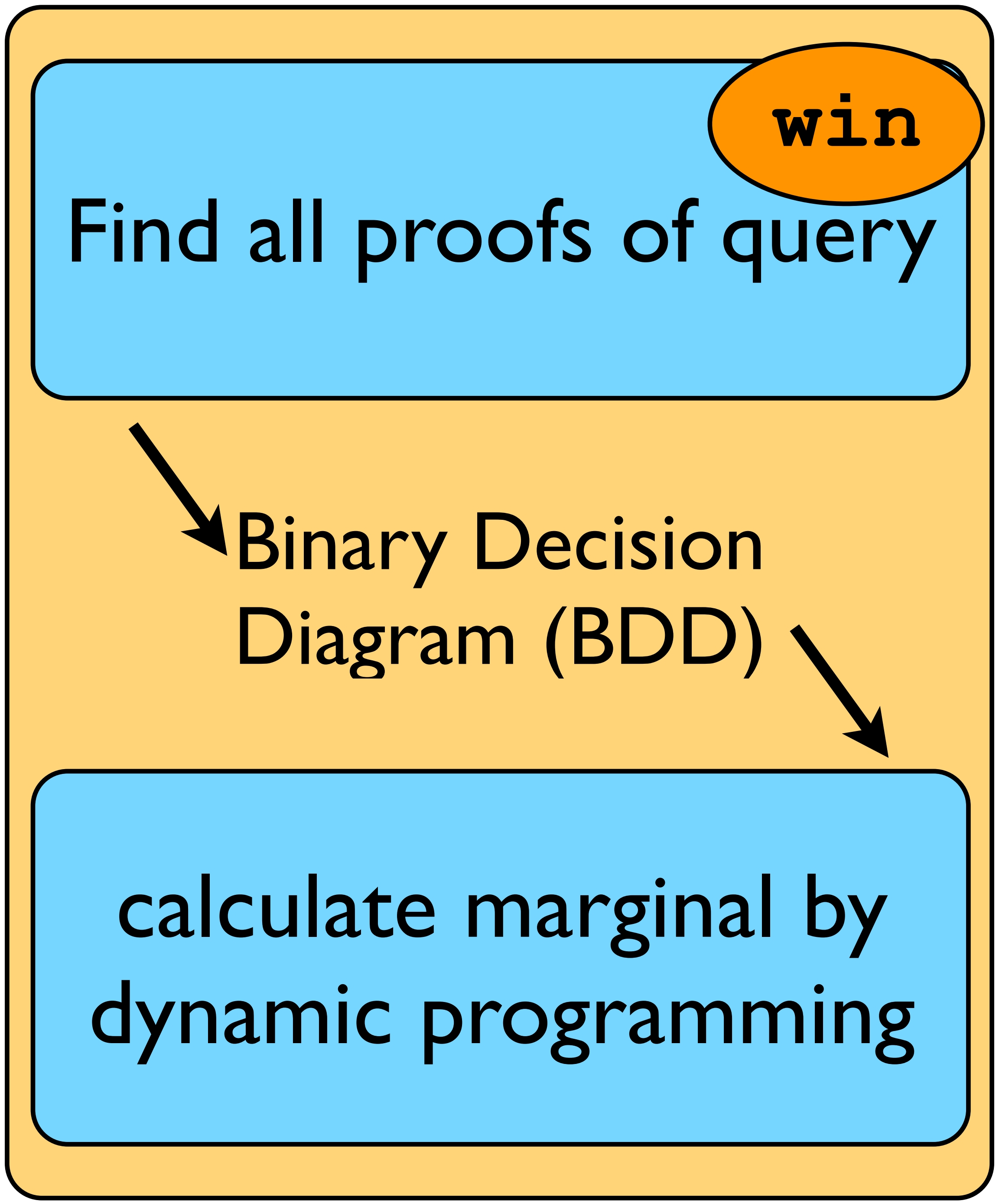
probability of `smokes(c)`?

`smokes(c) = i(b,c)&s(b) v i(b,c)&i(a,b)&s(a)`

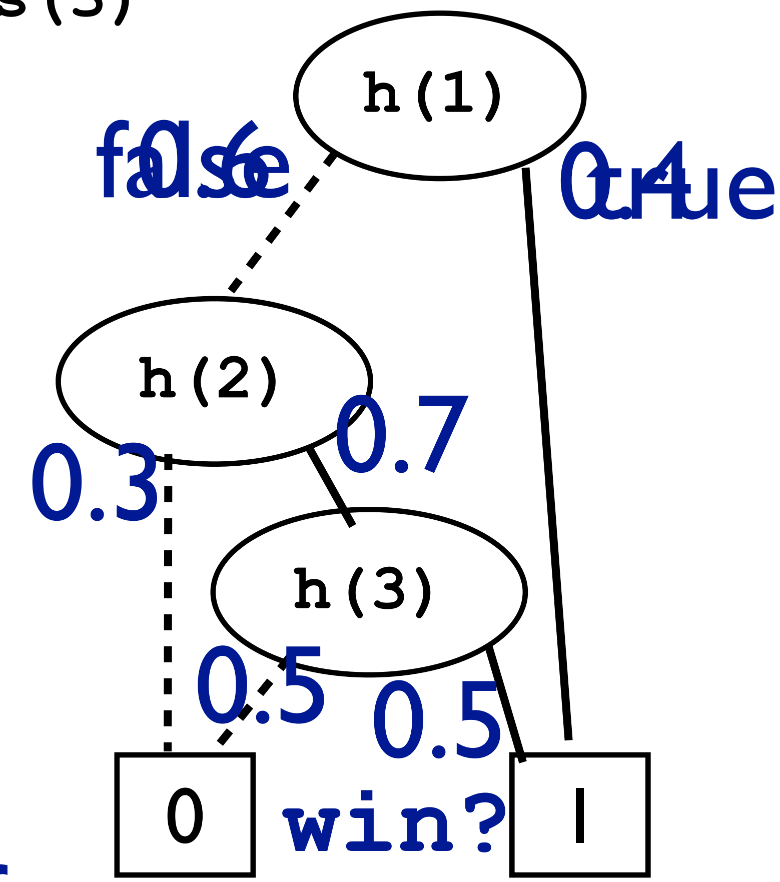
Initial Approach

(ProbLog1 & others)

```
0.4::heads(1).
0.7::heads(2).
0.5::heads(3).
win :- heads(1).
win :- heads(2),heads(3).
```



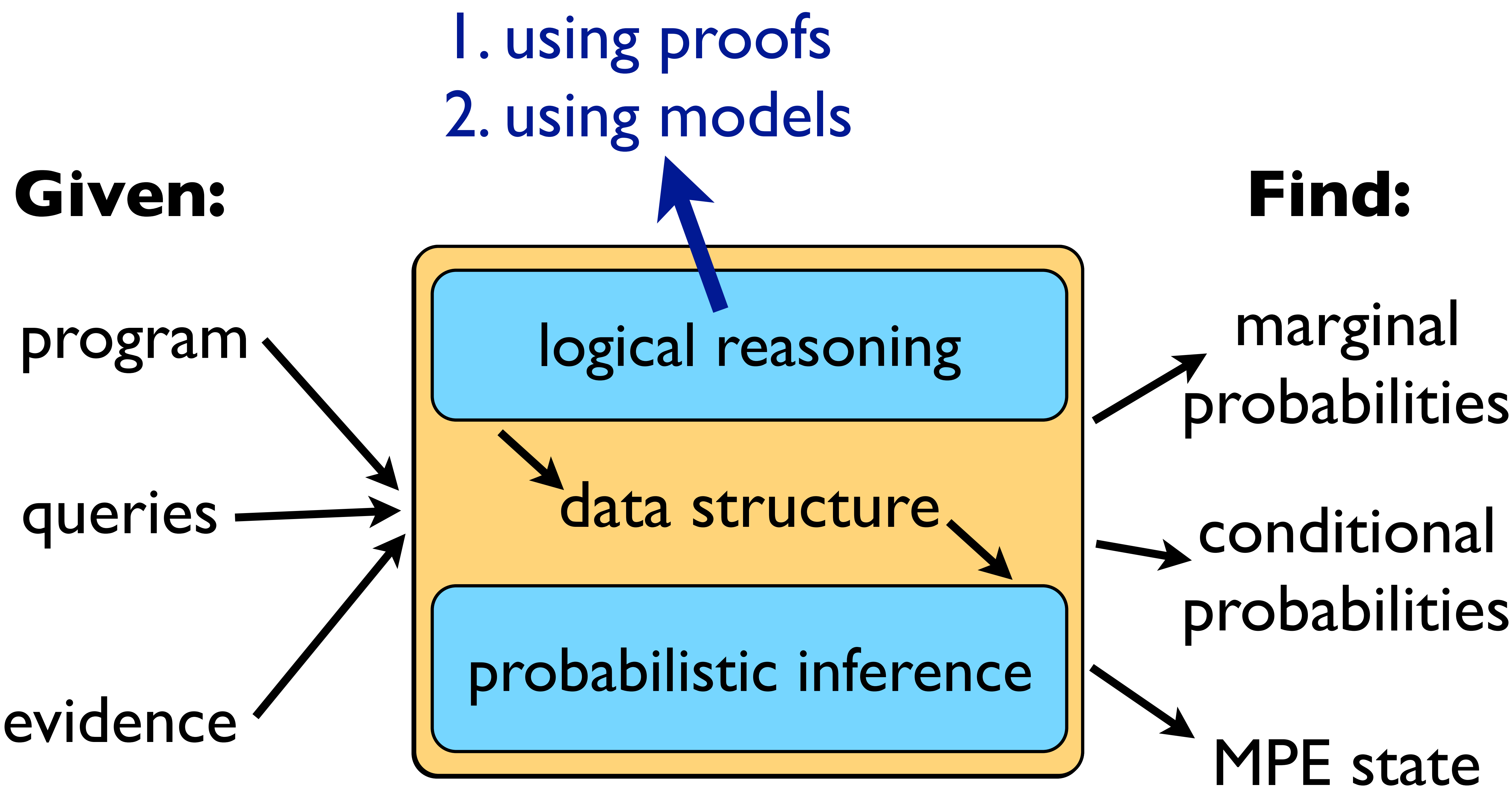

```
heads(1)
heads(2)  & heads(3)
```



P(**win**) = probability of reaching 1-leaf

[De Raedt et al, IJCAI 07; Kimmig et al, TPLP 11]

Answering Questions



Logical Reasoning: Models in Prolog

```
stress(ann).
influences(ann,bob).
influences(bob,carl).

smokes(X) :- stress(X).
smokes(X) :-
      influences(Y,X),
      smokes(Y).
```

```
?- smokes(carl).
```

- Forward reasoning to construct unique model:
  - Start with database facts
  - Use rules to add more facts

```
stress(ann).
influences(ann,bob).
influences(bob,carl).

smokes(ann).
smokes(bob).
smokes(carl).
```

- Query true iff in model
- **ProbLog**: each possible world is a model, probability of query is sum over models where query is true

→ weighted model counting

Weighte

$$P(Q) = \sum_{F \cup R \models Q} \prod_{f \in F} p(f) \prod_{f \not\in F} 1 - p(f)$$

propositional formula in conjunctive normal form (CNF)

given by ProbLog program & query

$$WMC(\phi) = \sum_{I_V \models \phi} \prod_{l \in I_V} w(l)$$

interpretations (truth value assignments) of propositional variables

possible worlds

weight of literal

for p::f,

w(f) = p

w(not f) = 1−p

ProbLog → CNF

```
0.8::stress(ann).
0.4::stress(bob).
0.6::influences(ann,bob).
0.2::influences(bob,carl).

smokes(X) :- stress(X).
smokes(X) :-
      influences(Y,X),
      smokes(Y).
```

```
?- smokes(carl).
```

- Find relevant ground rules by backward reasoning

```
smokes(carl) :- influences(bob,carl),smokes(bob).
smokes(bob) :- stress(bob).
smokes(bob) :- influences(ann,bob),smokes(ann).
smokes(ann) :- stress(ann).
```

- Convert to propositional logic formula

may require loop-breaking

$$\mathrm{sm}(c) \leftrightarrow (i(b,c) \wedge \mathrm{sm}(b))$$
$$\wedge\ \mathrm{sm}(b) \leftrightarrow (\mathrm{st}(b) \vee (i(a,b) \wedge \mathrm{sm}(a)))$$
$$\wedge\ \mathrm{sm}(a) \leftrightarrow \mathrm{st}(a)$$

- Rewrite in CNF (as usual)

Current Approach
(ProbLog2)

```
0.4::heads(1).
0.7::heads(2).
0.5::heads(3).
win :- heads(1).
win :- heads(2),
              heads(3).
```

Find relevant ground program for queries & evidence

win

Weighted CNF

use weighted model counting / satisfiability

```
win :- heads(1).
win :- heads(2), heads(3).
```

$$\text{win} \leftrightarrow h(1) \lor (h(2) \land h(3))$$

$$(\neg\text{win} \lor h(1) \lor h(2))$$
$$\land\ (\neg\text{win} \lor h(1) \lor h(3))$$
$$\land\ (\text{win} \lor \neg h(1))$$
$$\land\ (\text{win} \lor \neg h(2) \lor \neg h(3))$$

use standard tool

$h(1) \rightarrow 0.4$ $\quad h(2) \rightarrow 0.7$ $\quad h(3) \rightarrow 0.5$

$\neg h(1) \rightarrow 0.6$ $\quad \neg h(2) \rightarrow 0.3$ $\quad \neg h(3) \rightarrow 0.5$

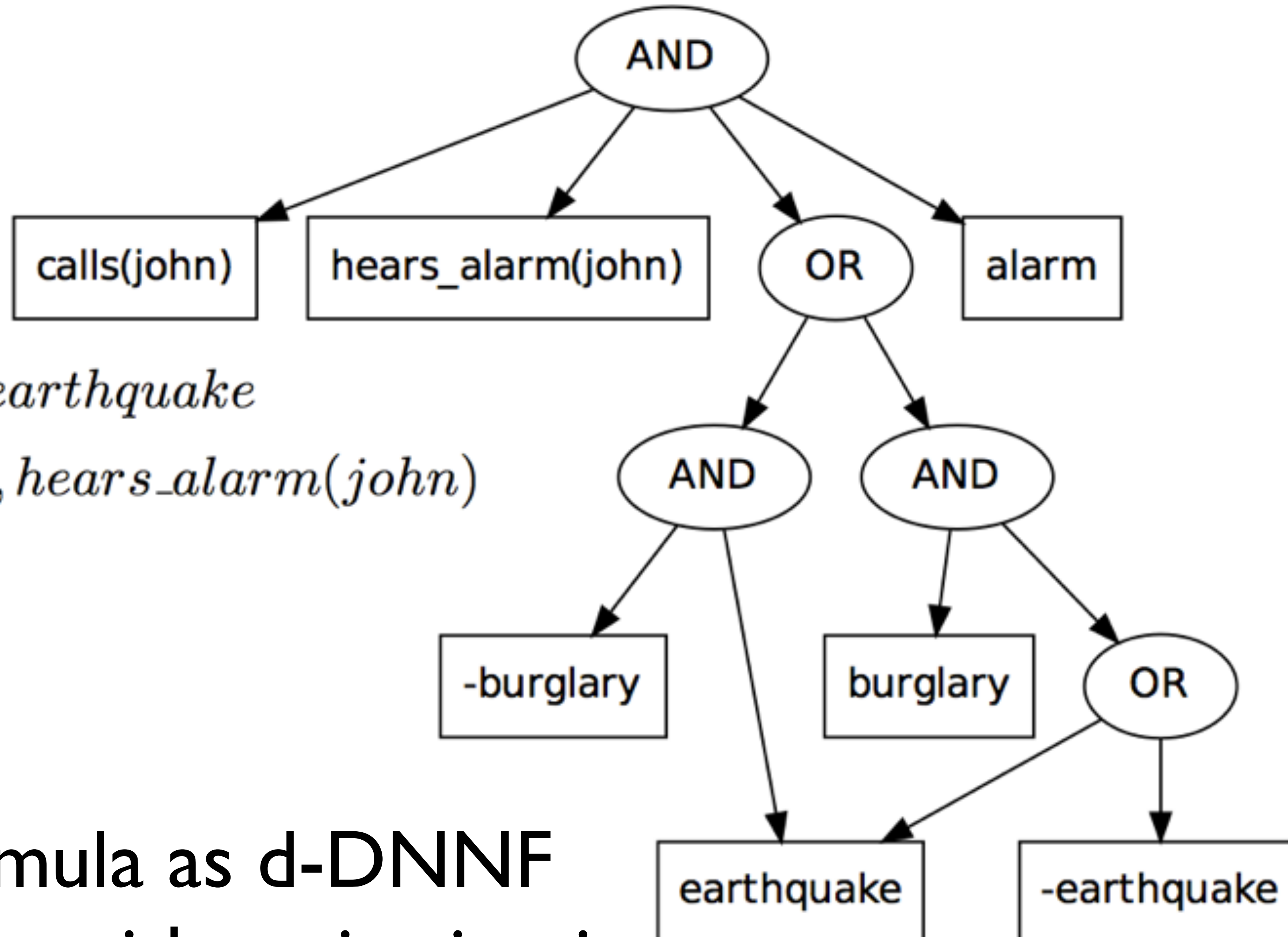
[Fierens et al, TPLP 14]

WMC using d-DNNFs

$$alarm \leftrightarrow burglary \vee earthquake$$

$$calls(john) \leftrightarrow alarm, hears\_alarm(john)$$

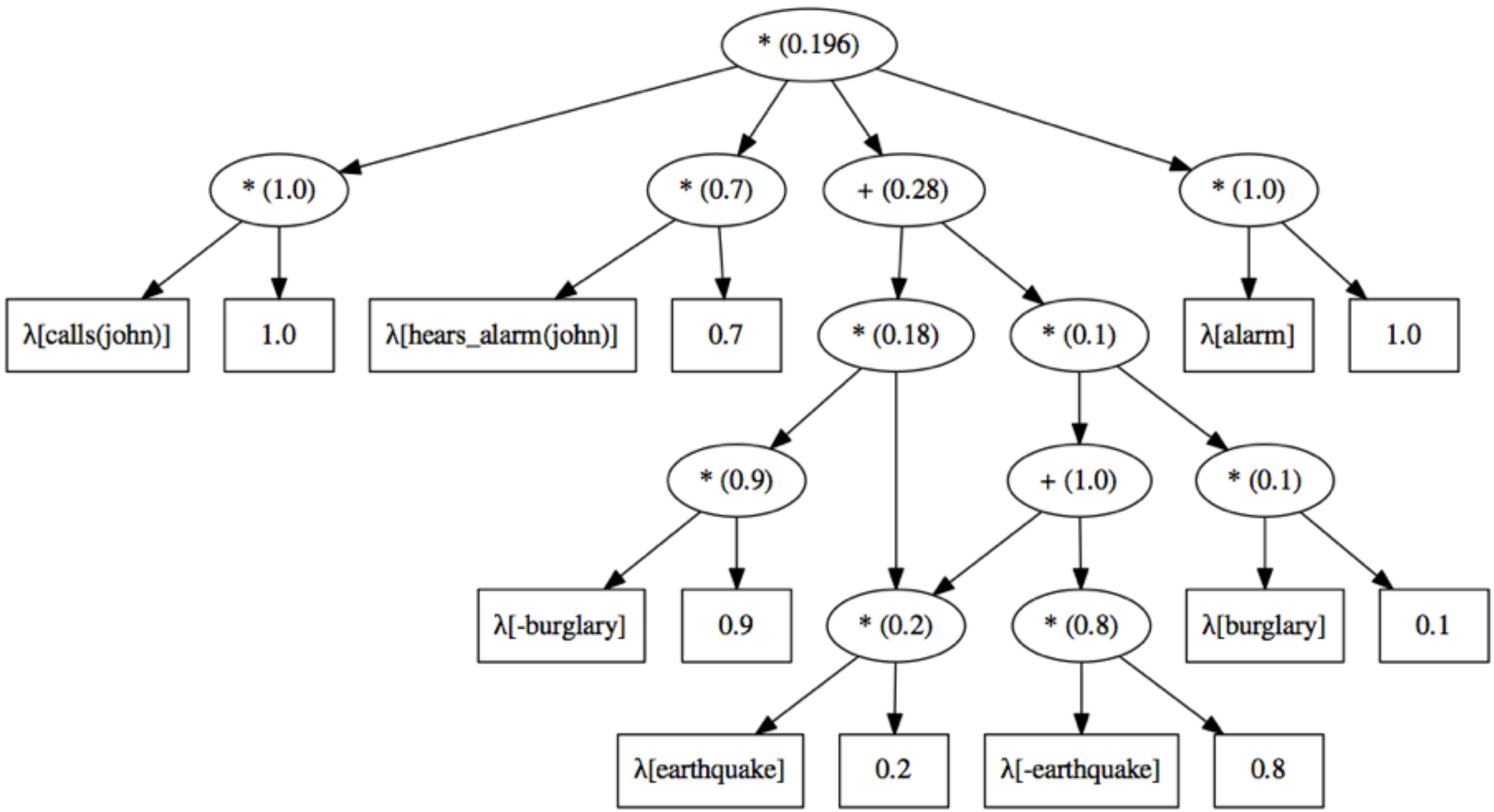
$$calls(john)$$



1. represent formula as d-DNNF
2. transform into arithmetic circuit
3. evaluate bottom-up

[Figure: Fierens et al, TPLP 14]

WMC using d-DNNFs



3. evaluate bottom-up

[Figure: Fierens et al, TPLP 14]

ProbLog Inference

- reduction to propositional formula
- addresses disjoint-sum-problem
- **but**: not all probabilistic logic programs face this problem! e.g., weather
- more generally: mutually exclusive proofs as assumed in PRISM
- more generally: a lot of work on approximate inference

Approximate Inference

- Lower and upper bounds

$$\phi_L \quad \models \quad \phi \quad \models \quad \phi_U$$

$$P(\phi_L) \ \leq \ P(\phi) \ \leq \ P(\phi_U)$$

- Sampling

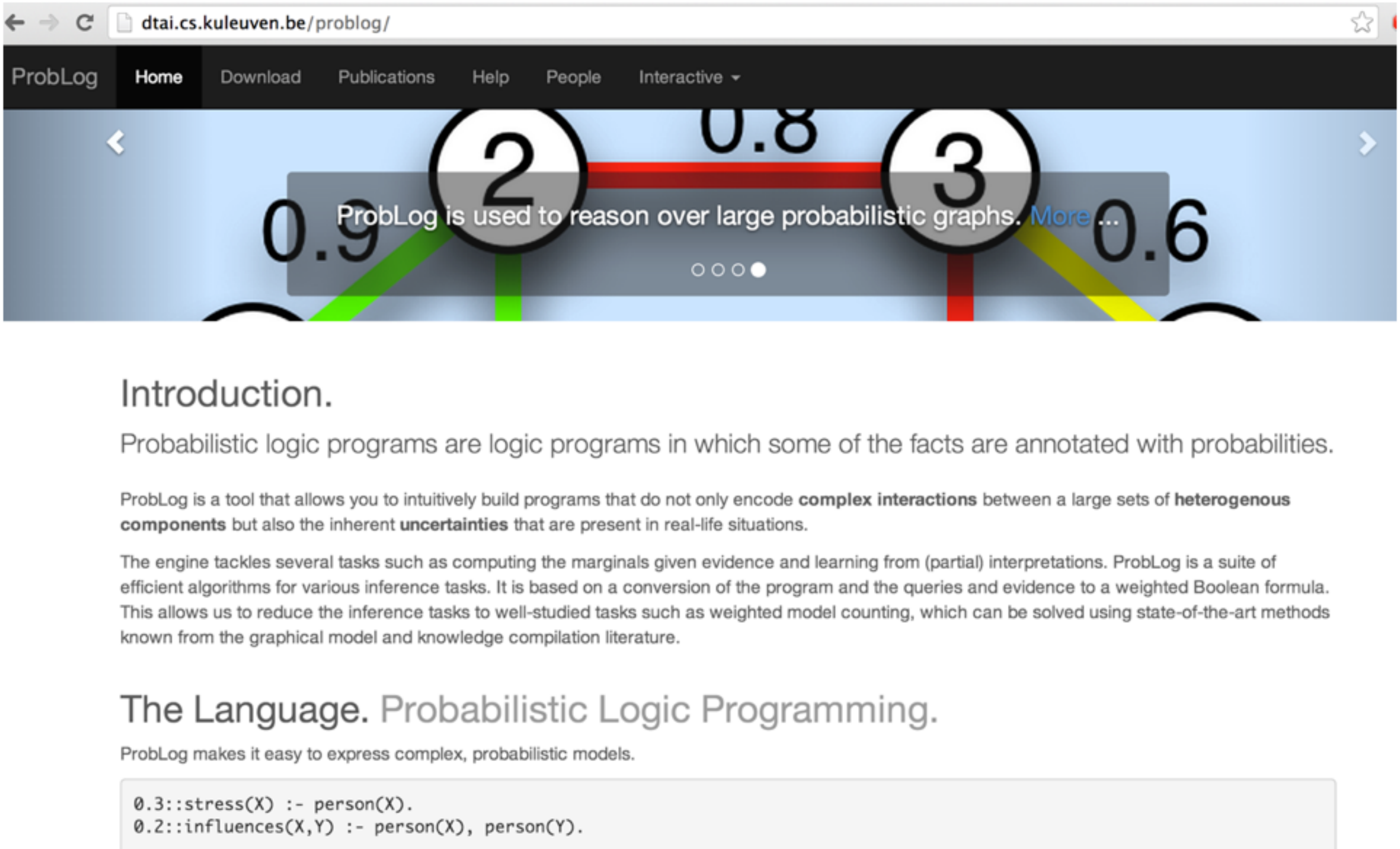
Parameter Learning

e.g., webpage classification model

for each *CLASS1*, *CLASS2* and each *WORD*

?? :: link_class(Source,Target,*CLASS1*,*CLASS2*).
?? :: word_class(*WORD*,*CLASS*).

```
class(Page,C)    :-   has_word(Page,W), word_class(W,C).

class(Page,C)          :-   links_to(OtherPage,Page),
     class(OtherPage,OtherClass),
                                 link_class(OtherPage,Page,OtherClass,C).
```



dtai.cs.kuleuven.be/problog/

ProbLog Home Download Publications Help People Interactive

ProbLog is used to reason over large probabilistic graphs. More ...

Introduction.

Probabilistic logic programs are logic programs in which some of the facts are annotated with probabilities.

ProbLog is a tool that allows you to intuitively build programs that do not only encode **complex interactions** between a large sets of **heterogenous components** but also the inherent **uncertainties** that are present in real-life situations.

The engine tackles several tasks such as computing the marginals given evidence and learning from (partial) interpretations. ProbLog is a suite of efficient algorithms for various inference tasks. It is based on a conversion of the program and the queries and evidence to a weighted Boolean formula. This allows us to reduce the inference tasks to well-studied tasks such as weighted model counting, which can be solved using state-of-the-art methods known from the graphical model and knowledge compilation literature.

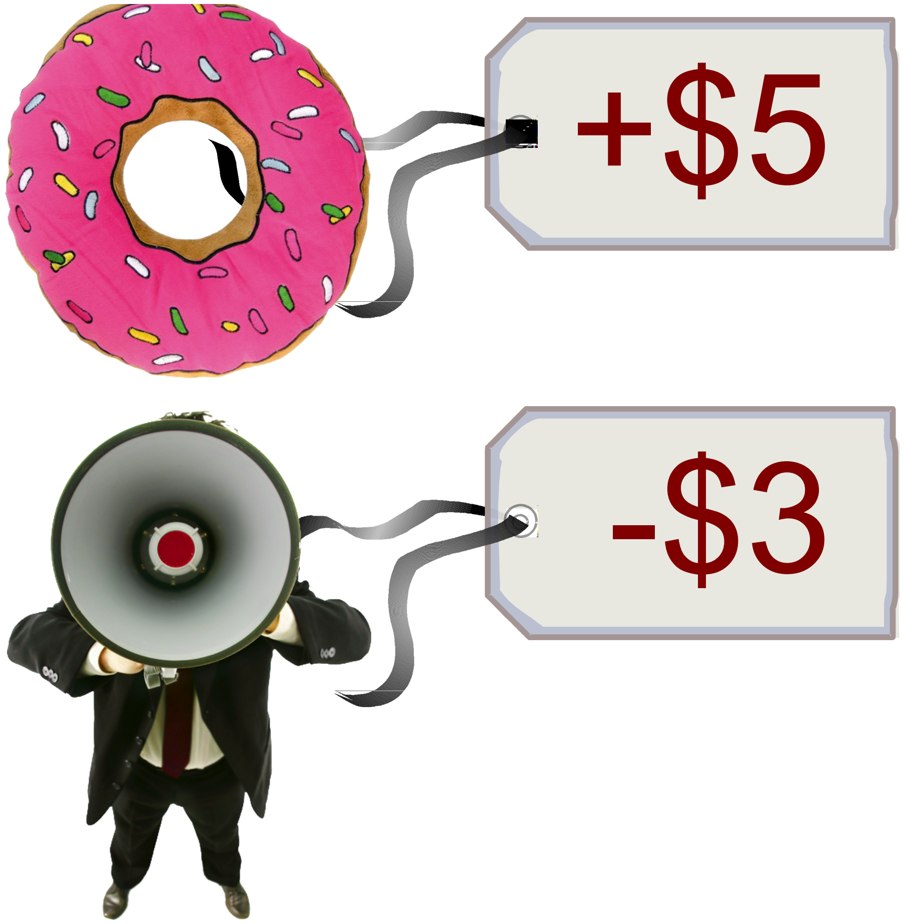
The Language. Probabilistic Logic Programming.

ProbLog makes it easy to express complex, probabilistic models.

```
0.3::stress(X) :- person(X).
0.2::influences(X,Y) :- person(X), person(Y).
```

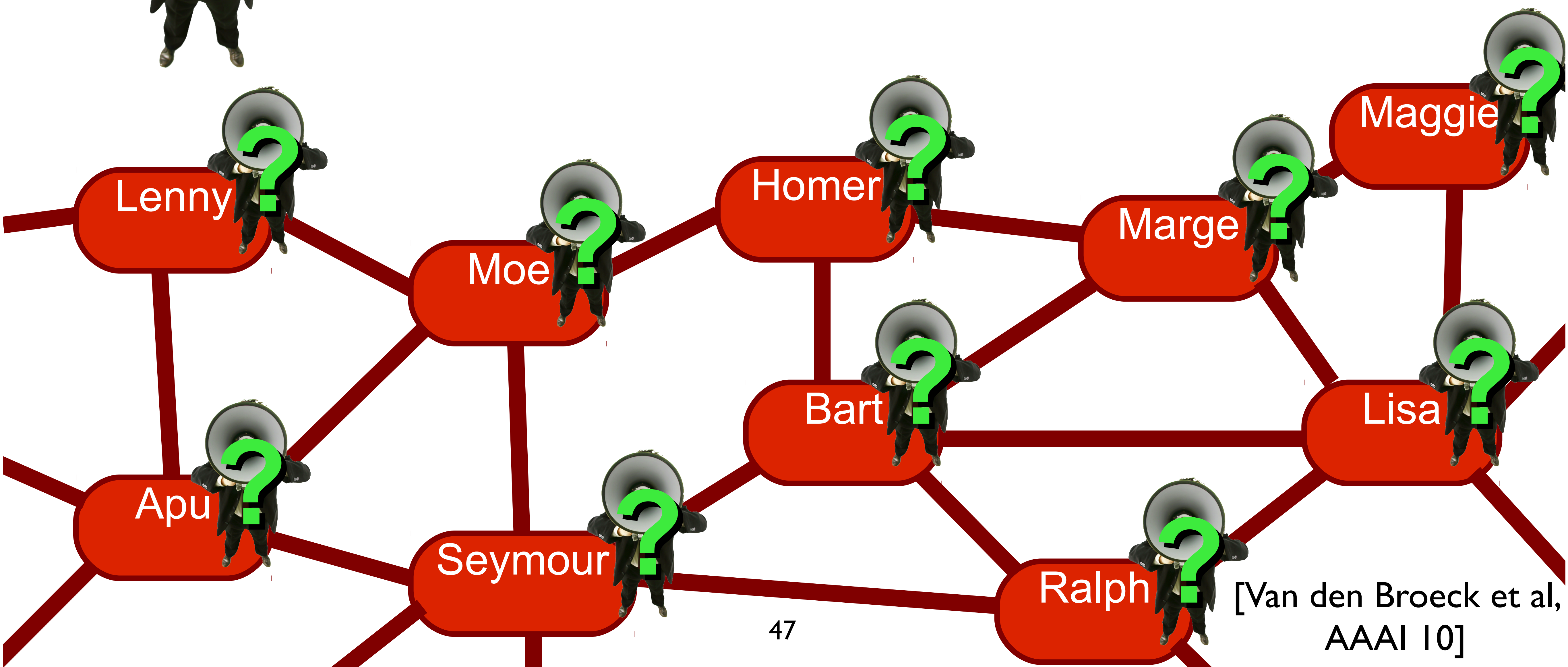
Part III: Decisions

Viral Marketing



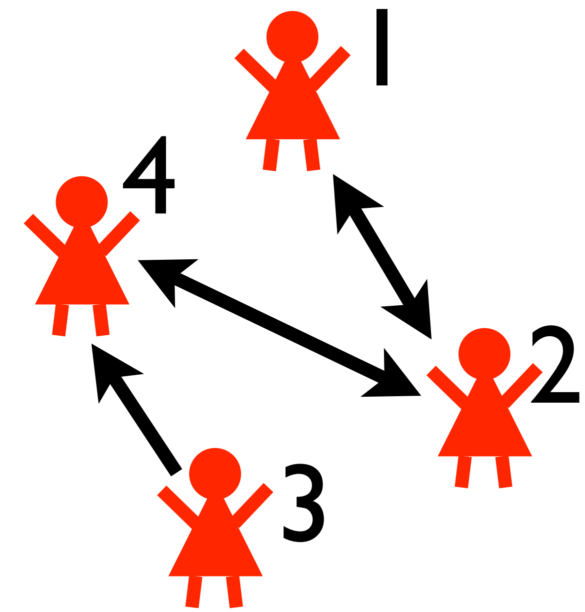
Which advertising strategy maximizes expected profit?

decide truth values of some atoms



[Van den Broeck et al, AAAI 10]

DTProbLog




```
? :: marketed(P) :- person(P).

0.3 :: buy_trust(X,Y) :- friend(X,Y).
0.2 :: buy_marketing(P) :- person(P).

buys(X) :- friend(X,Y), buys(Y), buy_trust(X,Y).
buys(X) :- marketed(X), buy_marketing(X).

buys(P) => 5 :- person(P).
marketed(P) => -3 :- person(P).
```

```
person(1).
person(2).
person(3).
person(4).

friend(1,2).
friend(2,1).
friend(2,4).
friend(3,4).
friend(4,2).
```

decision fact: true or false?

probabilistic facts + logical rules

utility facts: cost/reward if true

utility = −3 −3 + 5 + 5 = 4

probability = 0.0032

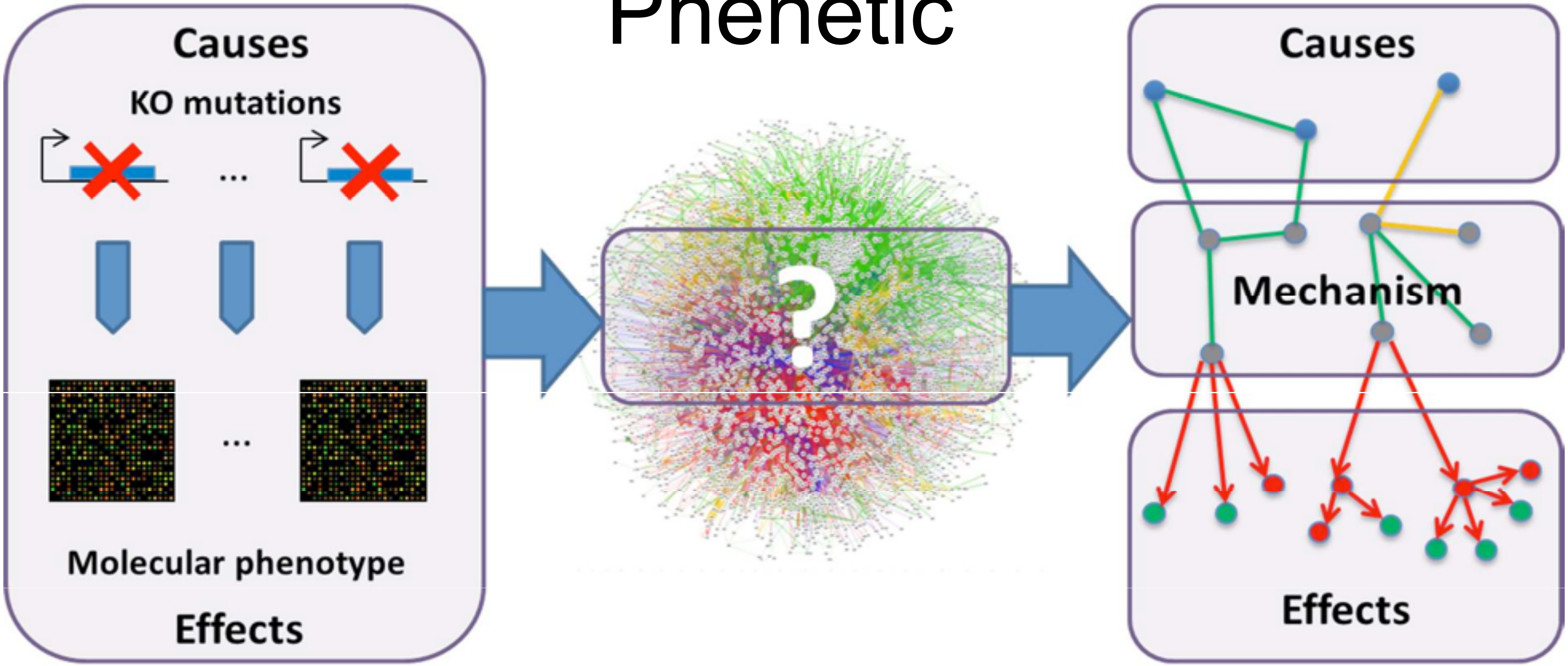
```
marketed(1)        marketed(3)

    bt(2,1)    bt(2,4)        bm(1)

buys(1)      buys(2)
```

task: find strategy that maximizes expected utility

solution: using ProbLog technology

world contributes 0.0032×4 to expected utility of strategy

Phenetic



- Causes: Mutations
  - All related to similar phenotype
- Effects: Differentially expressed genes
- 27 000 cause effect pairs

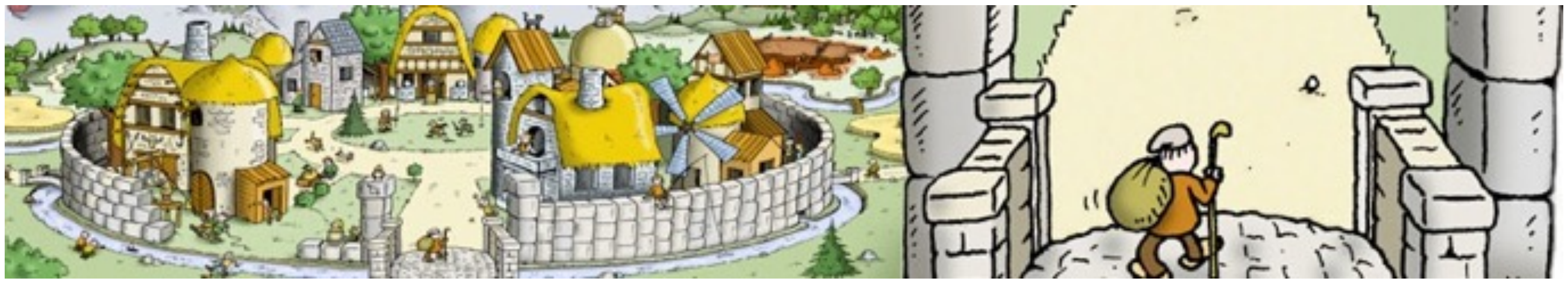
- Interaction network:
  - 3063 nodes
    - Genes
    - Proteins
  - 16794 edges
    - Molecular interactions
    - Uncertain

- Goal: connect causes to effects through common subnetwork
  - = Find mechanism
- Techniques:
  - DTProbLog
  - Approximate inference

[De Maeyer et al., Molecular Biosystems 13, NAR 15]

Part IV: Dynamics

Dynamics: Evolving Networks



- *Travian*: A massively multiplayer real-time strategy game
  - Commercial game run by TravianGames GmbH
  - ~3.000.000 players spread over different "worlds"
  - ~25.000 players in one world

[Thon et al., MLJ 11, ECML 08]

World Dynamics

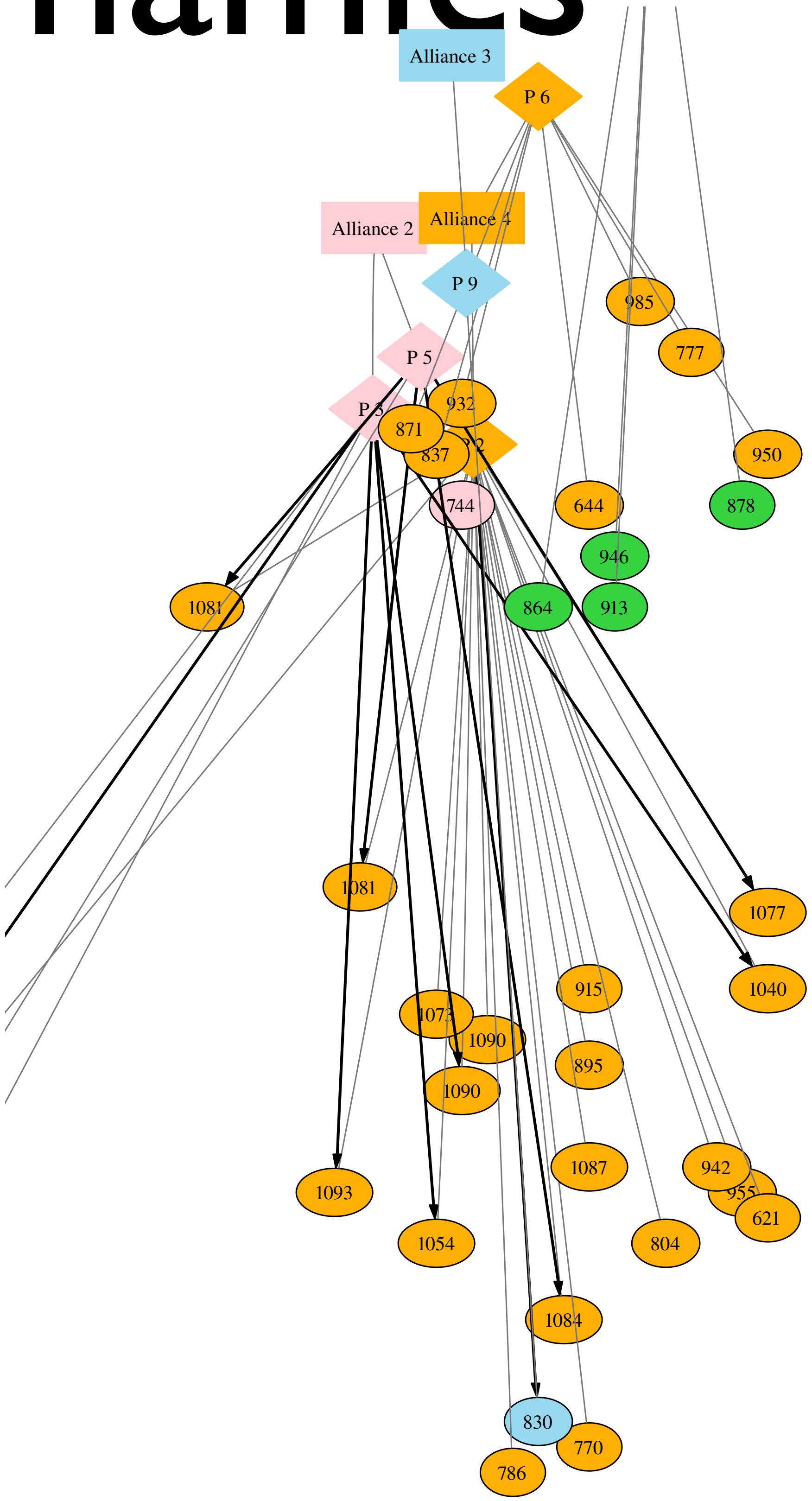
Fragment of world with

~10 alliances
~200 players
~600 cities

alliances color-coded

Can we build a model
of this world ?
Can we use it for playing
better ?

[Thon, Landwehr, De Raedt, ECML08]



World Dynamics

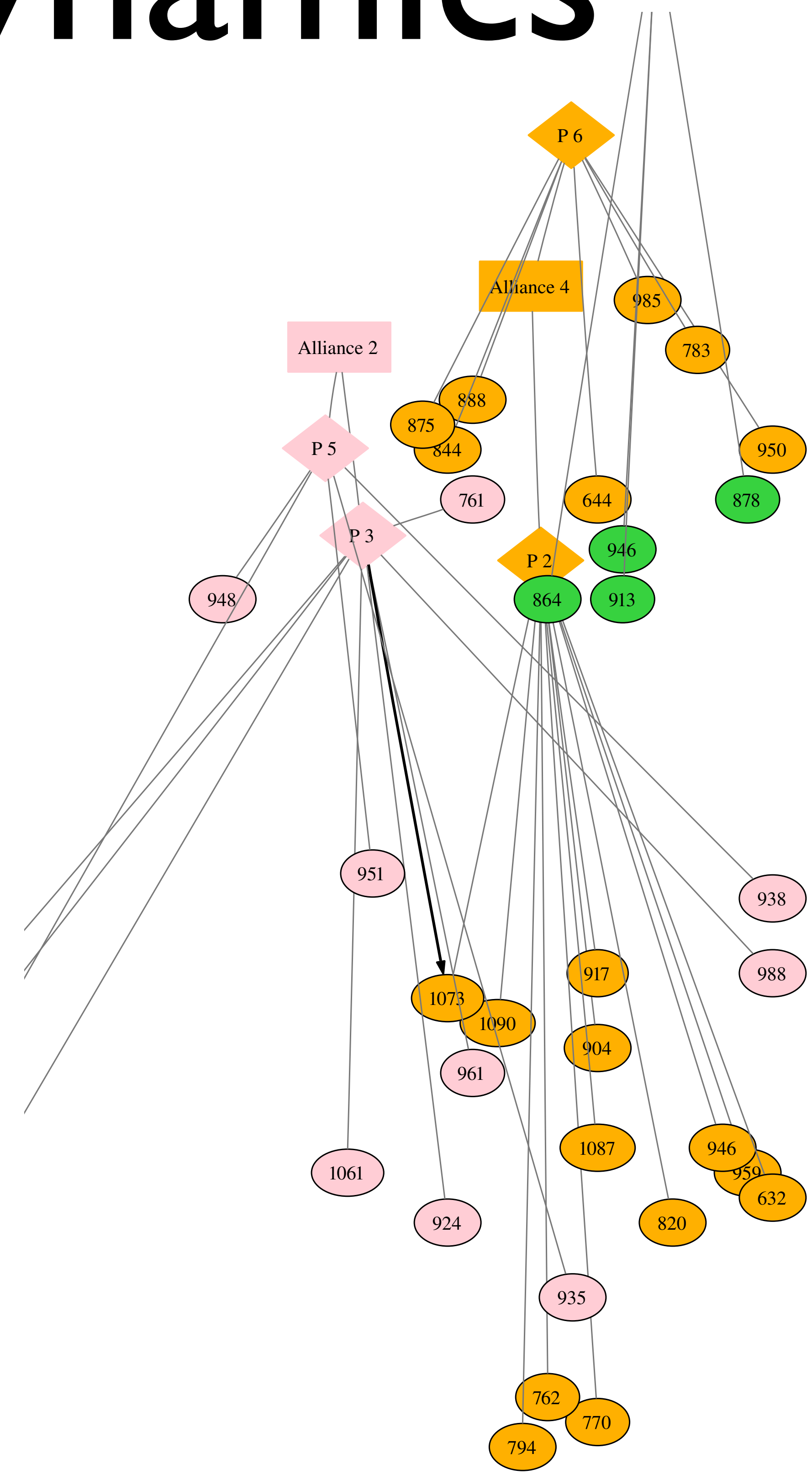
Fragment of world with

~10 alliances
~200 players
~600 cities

alliances color-coded

Can we build a model
of this world ?
Can we use it for playing
better ?

[Thon, Landwehr, De Raedt, ECML08]



World Dynamics

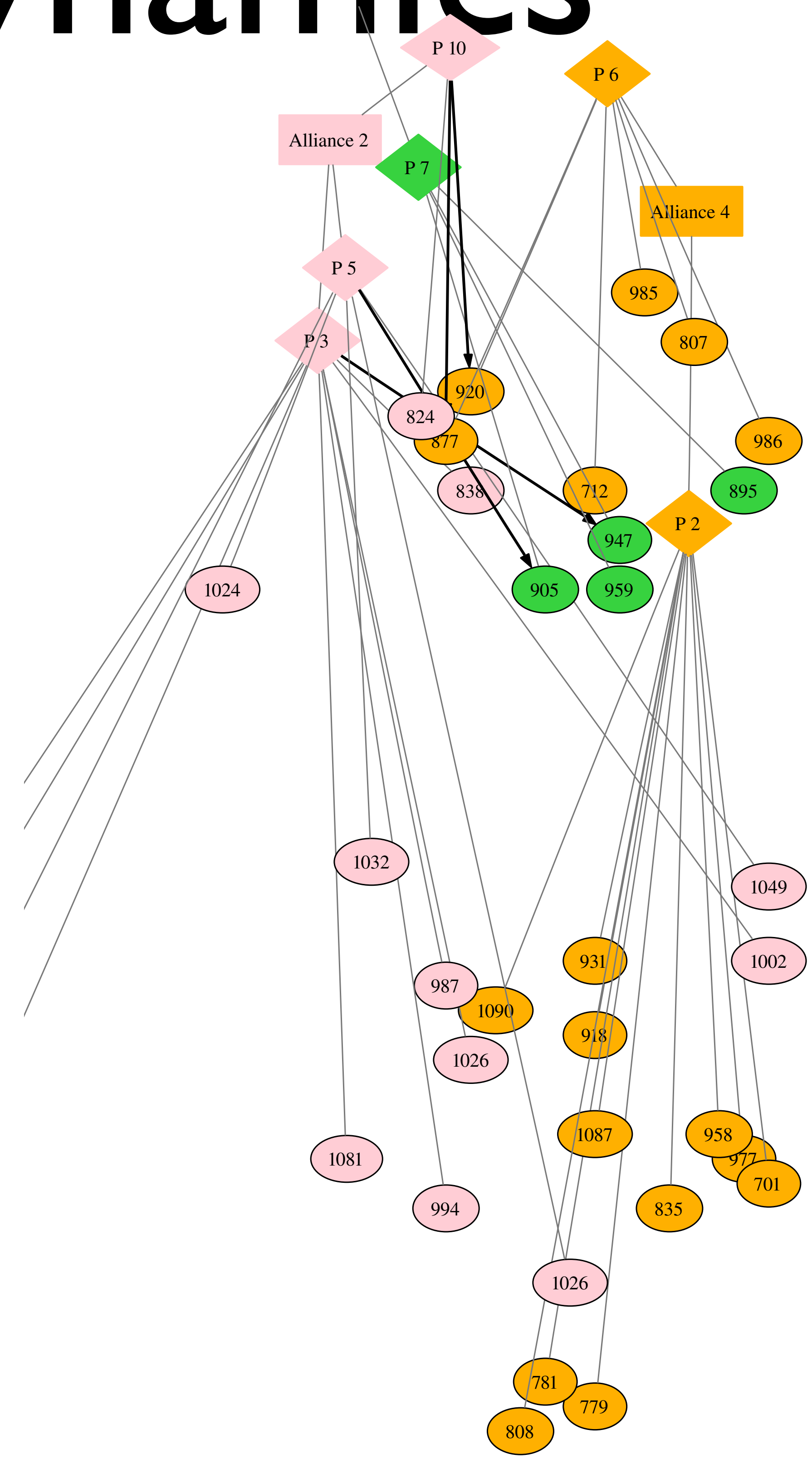
Fragment of world with

~10 alliances
~200 players
~600 cities

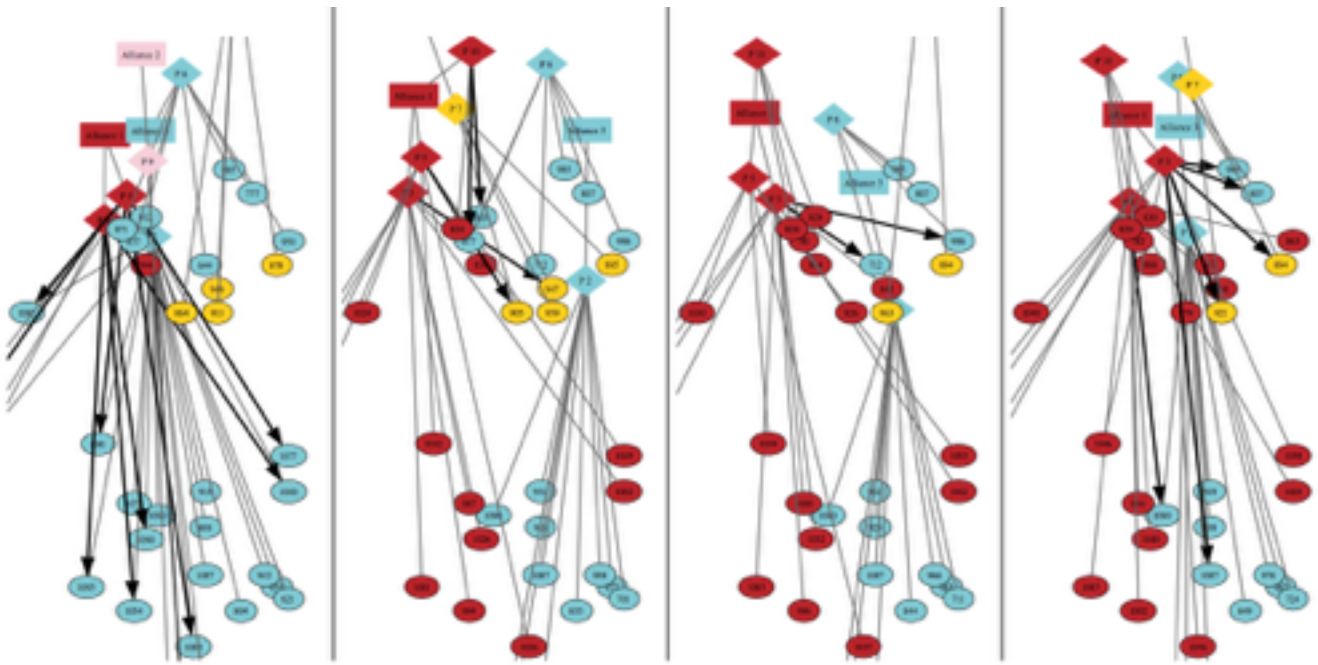
alliances color-coded

Can we build a model
of this world ?
Can we use it for playing
better ?

[Thon, Landwehr, De Raedt, ECML08]



Causal Probabilistic Time-Logic (CPT-L)



how does the world change over time?

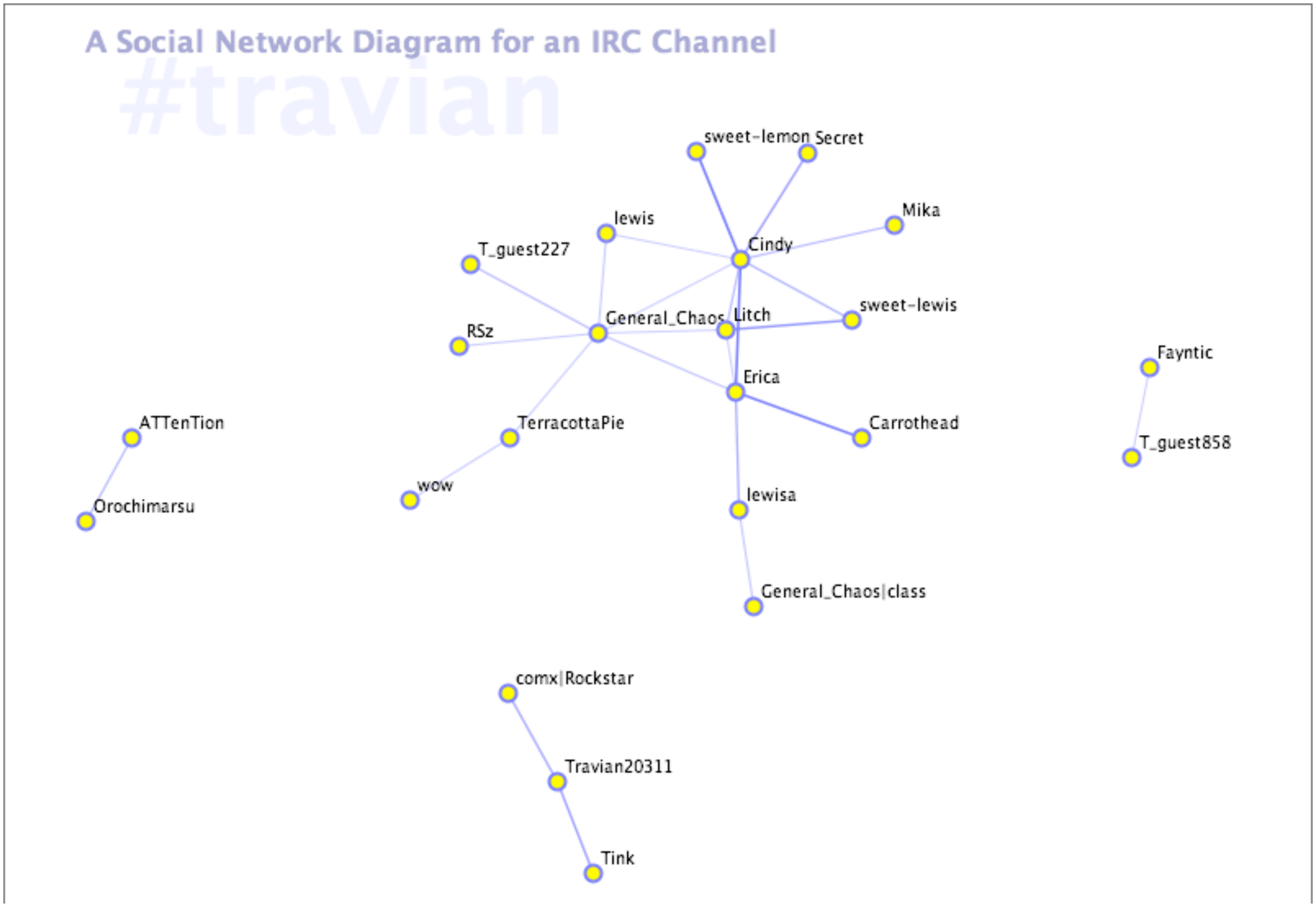
one of the **effects** holds at time T+1

```
0.4::conquest(Attacker,C); 0.6::nil <-
        city(C,Owner),city(C2,Attacker),close(C,C2).
```

if **cause** holds at time T

[Thon et al, MLJ 11]

Social Network of Chats



Limitations CPT-L

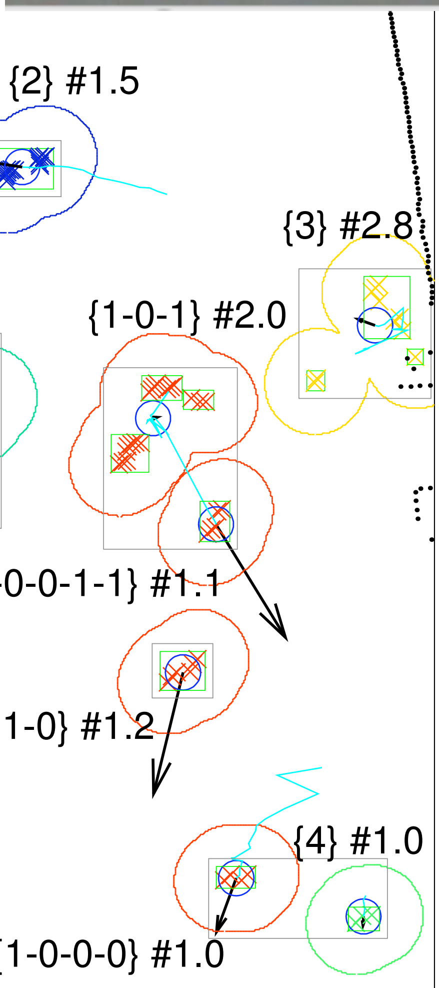
Inference slow / scalability

- uses knowledge compilation method
- compile formula for $P(I_{t+1}|I_{[0,t]})$
- exponential in number of time steps

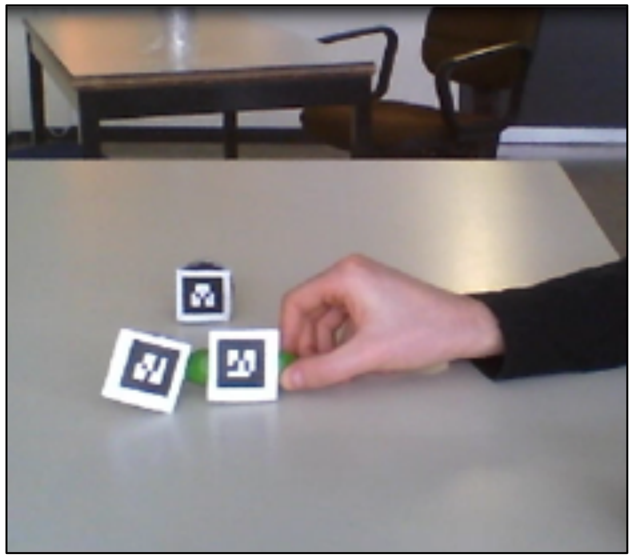
No continuous distributions

- needed for robotics / relational tracking applications

Relational Tracking



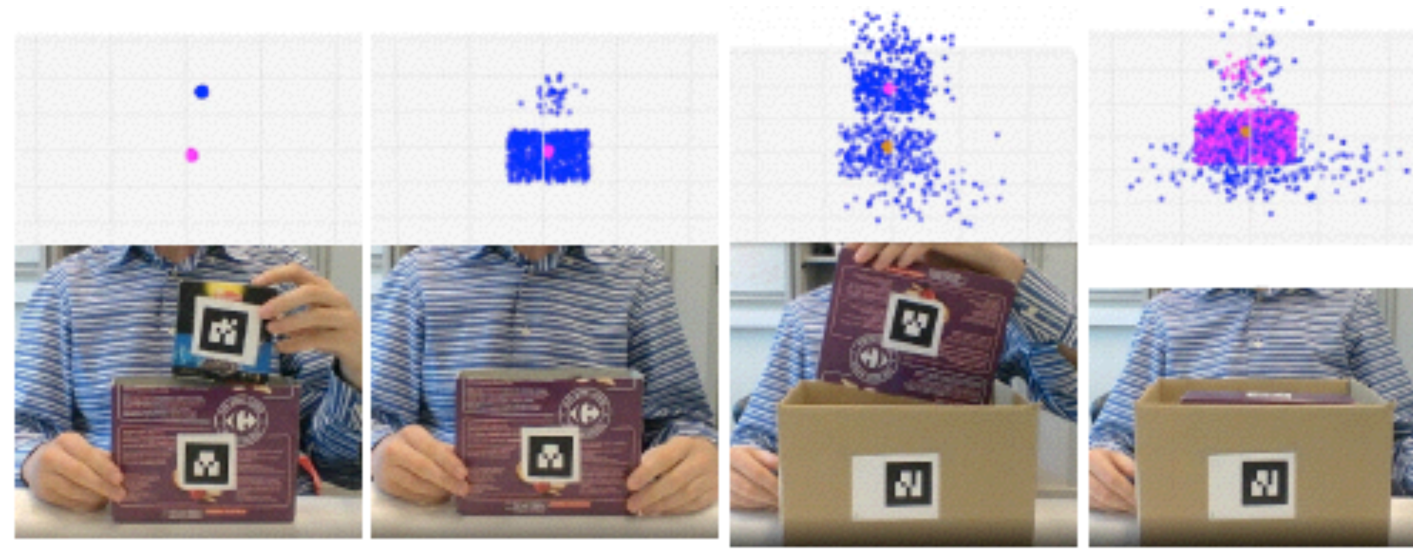
- Track people or objects over time? Even if temporarily hidden?
- Recognize activities?
- Infer object properties?



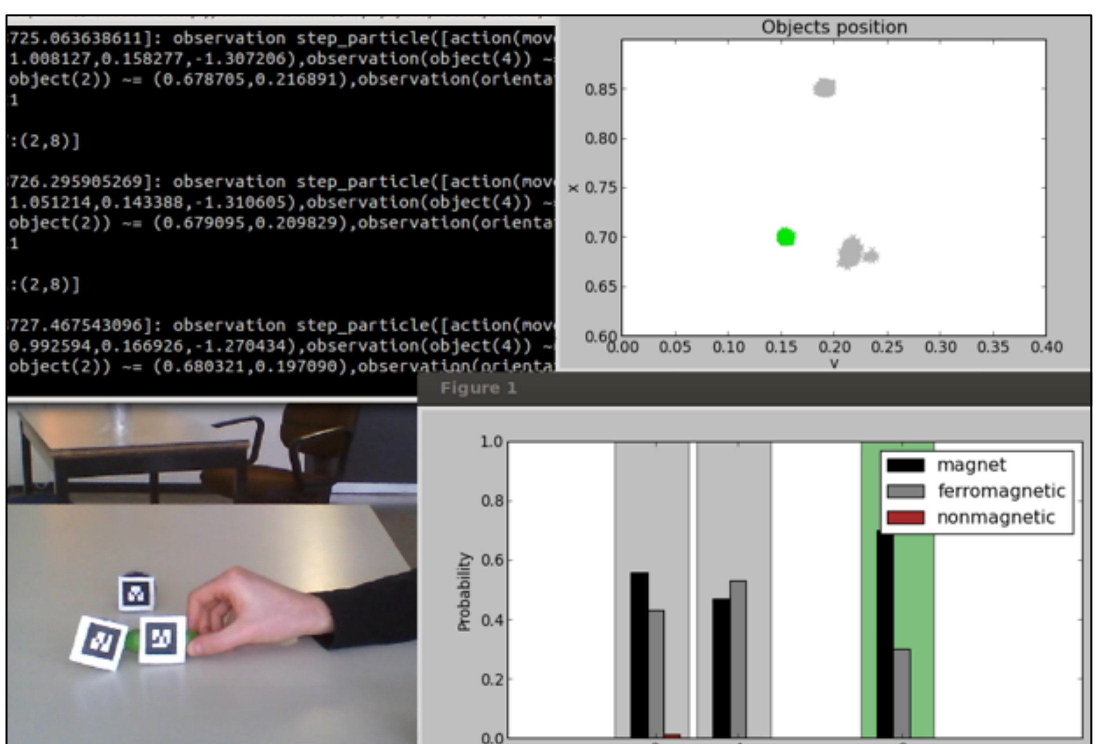
Relational State Estimation over Time

Magnetism scenario

- object tracking
- category estimation from interactions



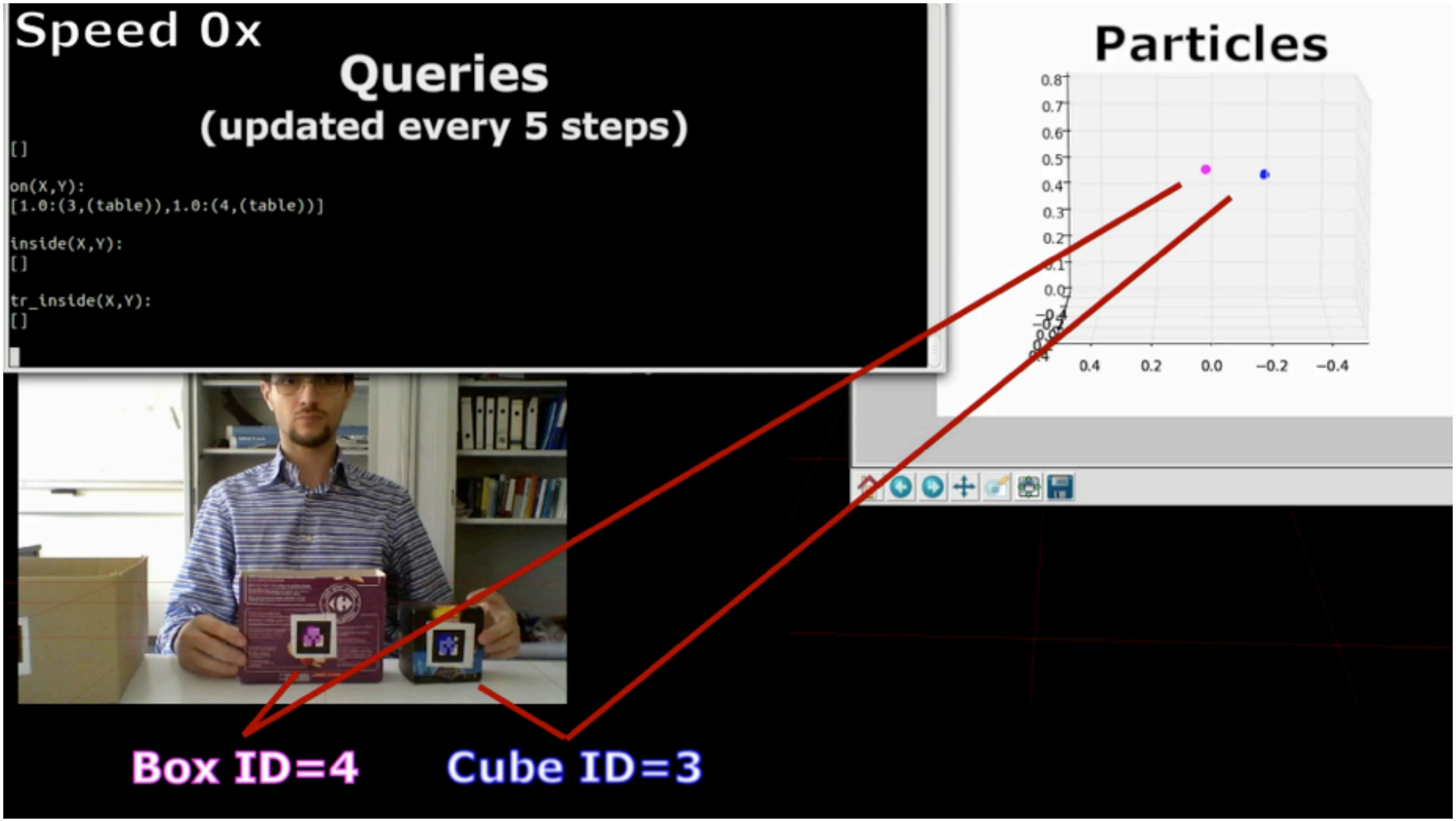
(a) cube on the box (b) cube inside the box (c) rotated box on a beige box (d) cube and box inside the beige box



Box scenario

- object tracking even when invisible
- estimate spatial relations

[Nitti et al, IROS 13]

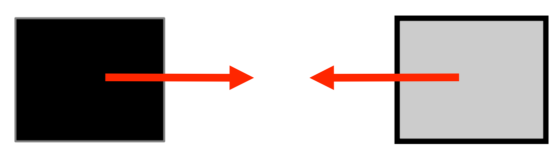

Speed 0x
Queries
(updated every 5 steps)
[]
on(X,Y):
[1.0:(3,(table)),1.0:(4,(table))]
inside(X,Y):
[]
tr_inside(X,Y):
[]
Particles
Box ID=4
Cube ID=3

IROS 13

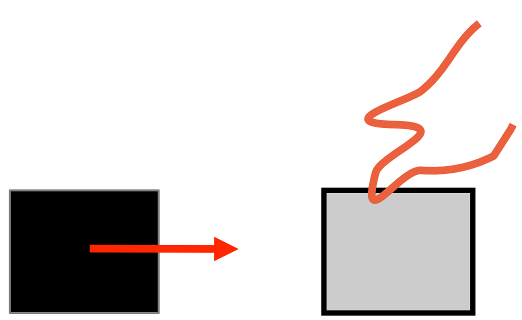
Magnetic scenario

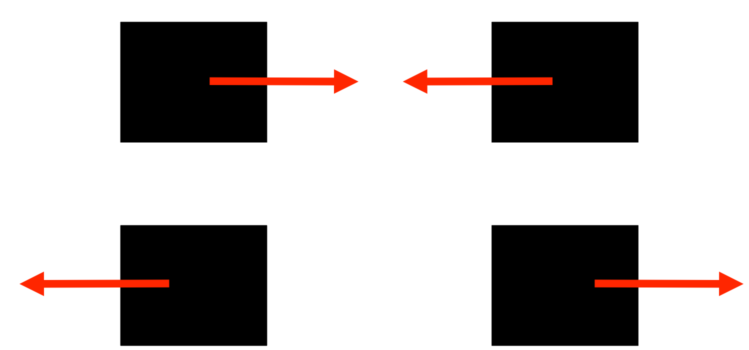
- 3 object types: magnetic, ferromagnetic, nonmagnetic

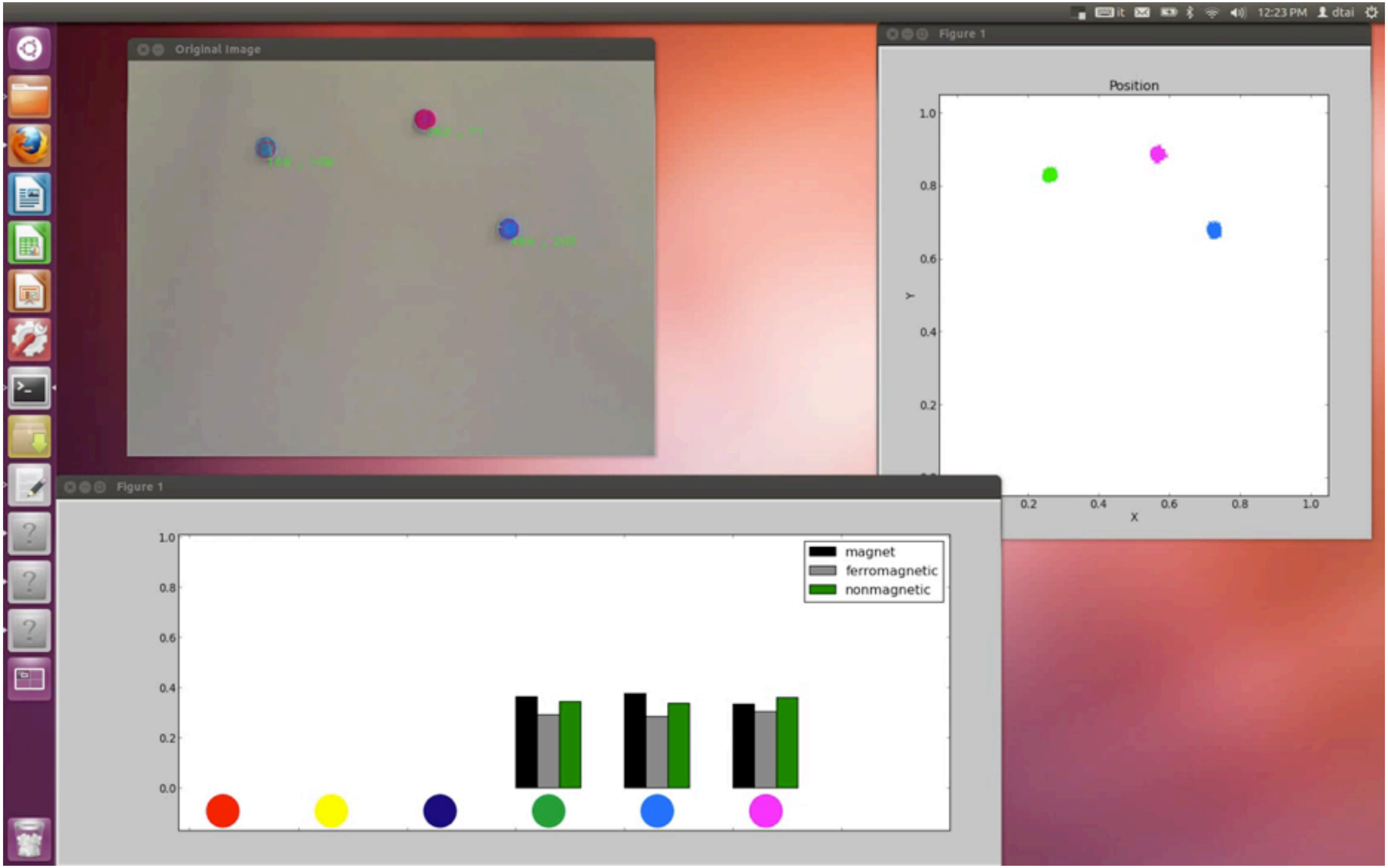
- Nonmagnetic objects do not interact
- A magnet and a ferromagnetic object attract each other



- Magnetic force that depends on the distance
- If an object is held magnetic force is compensated.







Original Image
Figure 1
Position
1.0
0.8
0.6
0.4
0.2
Y
0.2
0.4
0.6
0.8
1.0
X
Figure 1
1.0
0.8
0.6
0.4
0.2
0.0
magnet
ferromagnetic
nonmagnetic
12:23 PM
dtai

Magnetic scenario

- 3 object types: magnetic, ferromagnetic, nonmagnetic

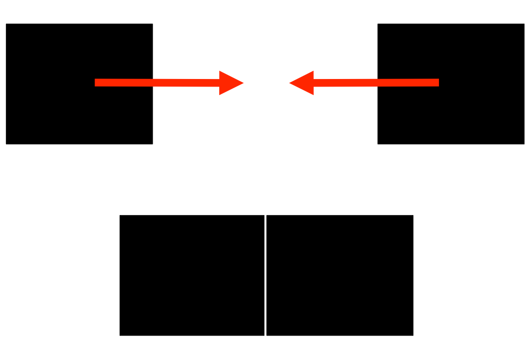
type(X)$_t$ ~ finite([1/3:magnet,1/3:ferromagnetic,1/3:nonmagnetic]) ← object(X).

- 2 magnets attract or repulse

interaction(A,B)$_t$ ~ finite([0.5:attraction,0.5:repulsion]) ← object(A), object(B), A<B,type(A)$_t$ = magnet,type(B)$_t$ = magnet.

- Next position after attraction

pos(A)$_{t+1}$ ~ gaussian(middlepoint(A,B)$_t$,Cov) ← near(A,B)$_t$, not(held(A)), not(held(B)), interaction(A,B)$_t$ = attr, c/dist(A,B)$_t^2$ > friction(A)$_t$.



pos(A)$_{t+1}$ ~ gaussian(pos(A)$_t$,Cov) ← not(attraction(A,B)).

Optimized inference: partial state

Distributional Clauses Particle Filter (DCPF)

Sampled

Pos(1)=(0, 3)
Pos(2)=(0, 1)

Marginalized

right(X,Y)
near(X,Y)
interaction(X,Y)
type(X) ~ [1/3:magnet,...]
[...]

Pos(1)=(0, 2)
Pos(2)=(0, 1)
near(1,2)=true
type(1)=nonmagnetic

right(X,Y)
near(X,Y)
interaction(X,Y)
type(X) ~ [1/3:magnet,...]
[...]

Classical particle filter

Pos(1)=(0, 3)
Pos(2)=(0, 1)
near(1,2)=false
near(2,1)=false
interaction(1,2)=none
type(1)=nonmagnetic
type(2)=nonmagnetic
[...]

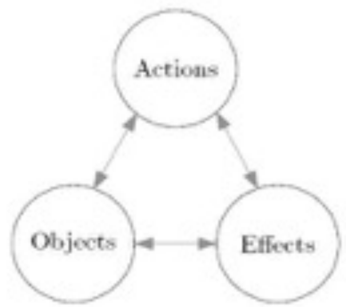
Pos(1)=(0, 2)
Pos(2)=(0, 1)
near(1,2)=true
near(2,1)=true
interaction(1,2)=none
type(1)=nonmagnetic
type(2)=nonmagnetic
[...]

Ongoing Work

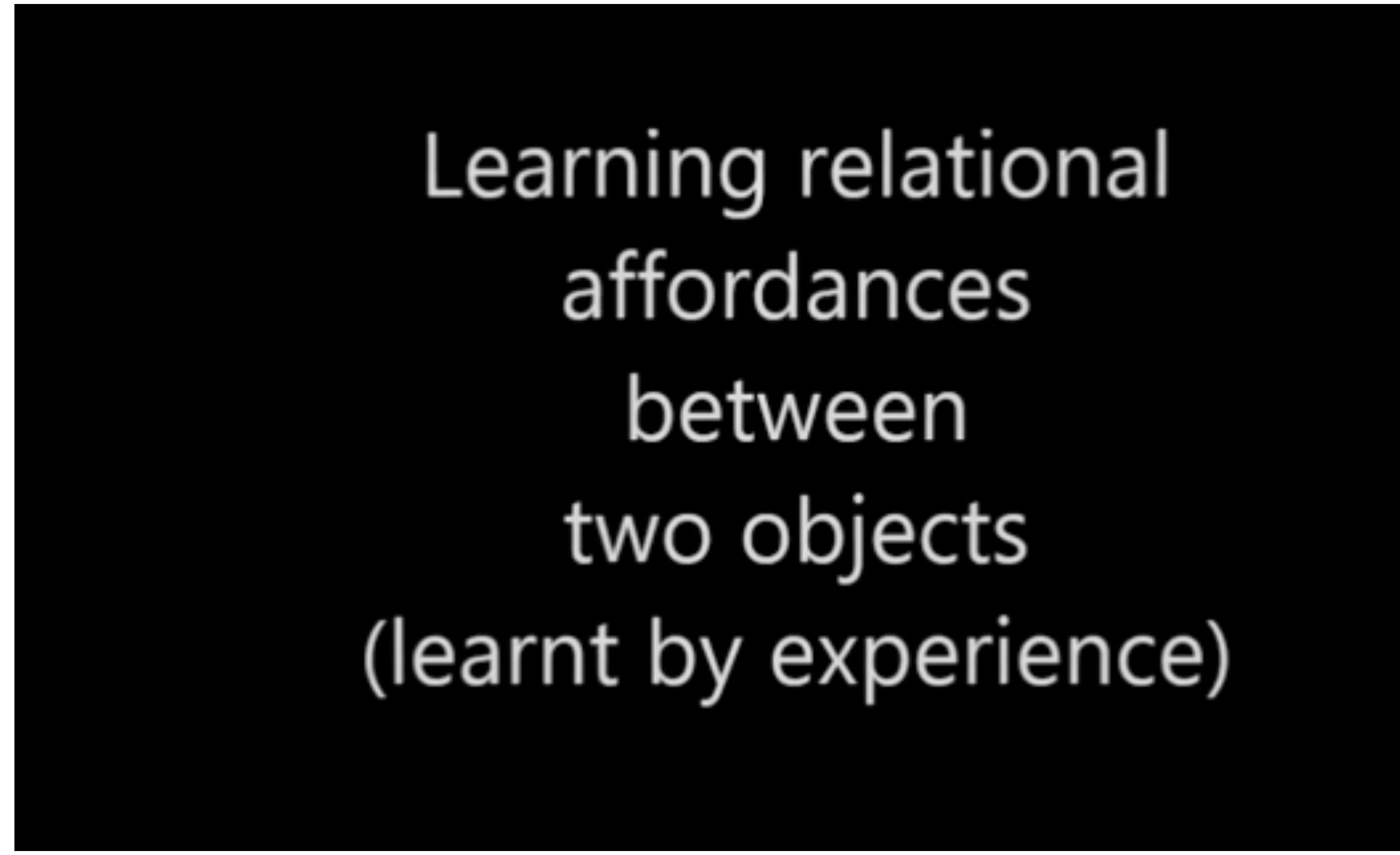
- Online parameter learning [Nitti, ICRA 2014]
- Integrate with planning
- Larger Experiments
- Applications in robotics (also to learn affordances)

Learning relational affordances

Learn probabilistic model



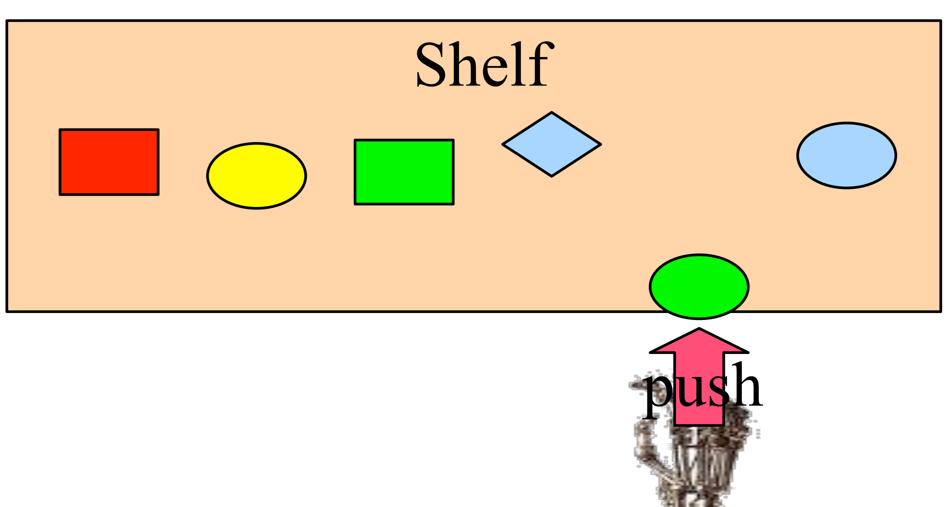
Inputs	Outputs	Function
(O, A)	E	Effect prediction
(O, E)	A	Action recognition/planning
(A, E)	O	Object recognition/selection

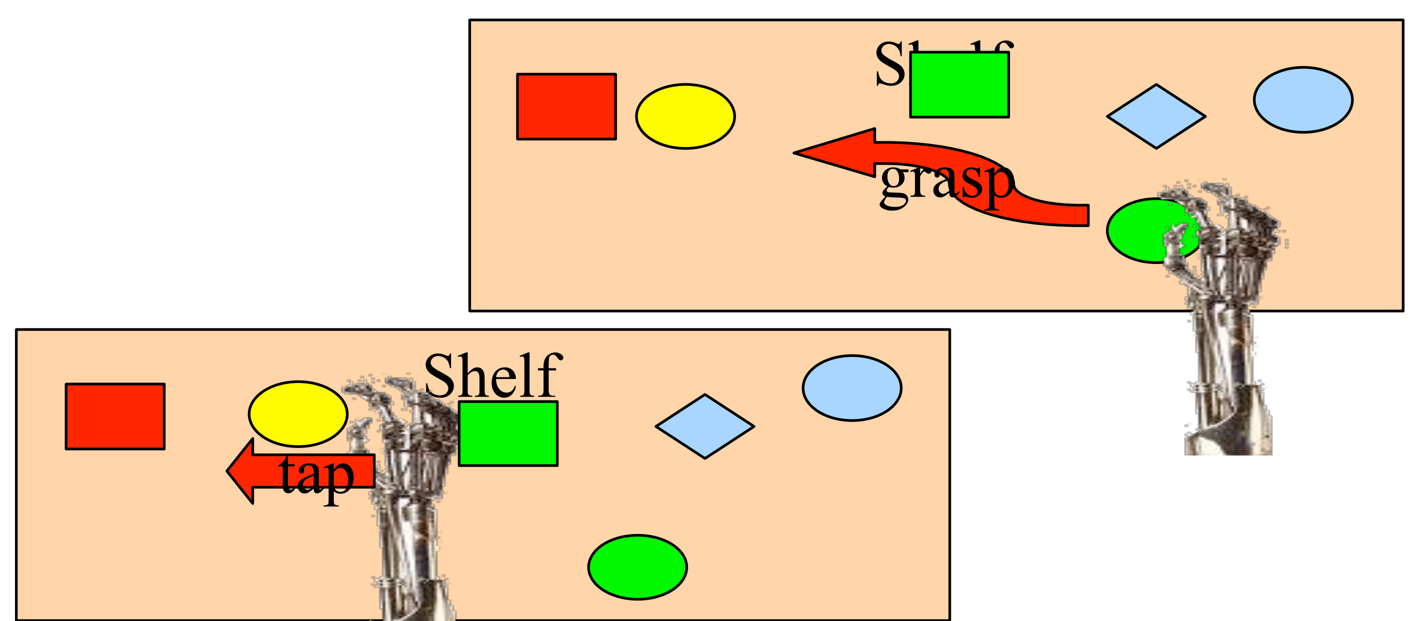


From two object interactions

Generalize to N

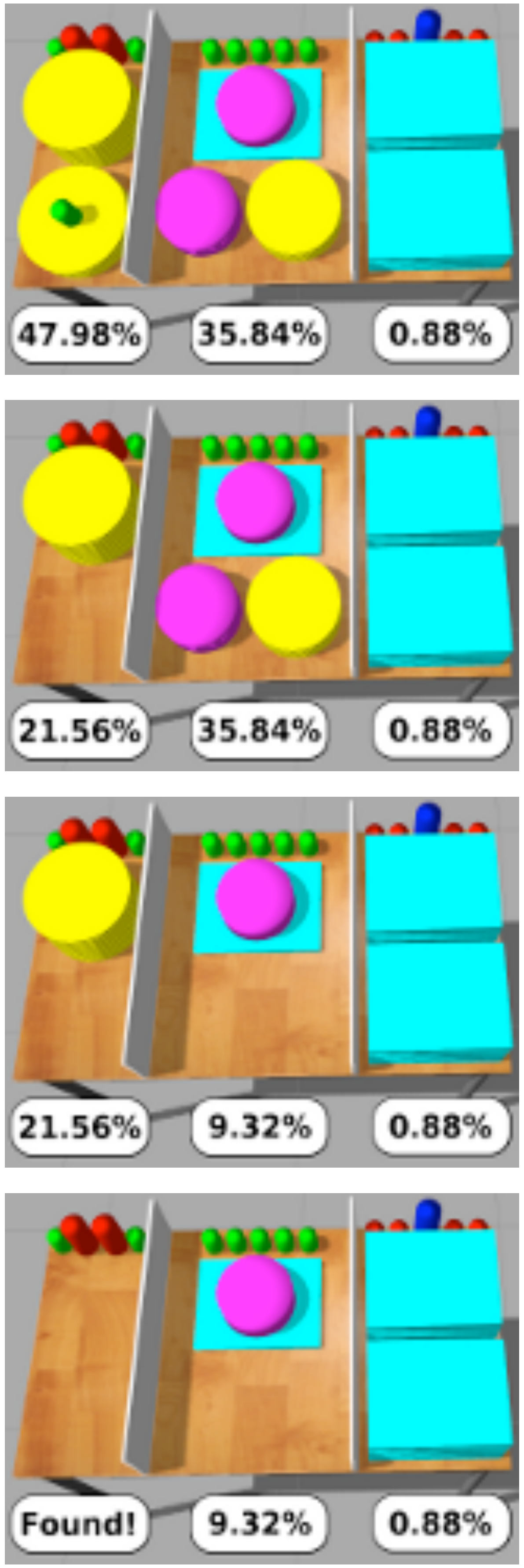
Moldovan et al. ICRA 12, 13, 14





Occluded Object Search

- How to achieve a specific configuration of objects on the shelf?
- Where's the orange mug?
- Where's something to serve soup in?
- Models of objects and their spatial arrangement



[Moldovan *et al.* 14]

ProbLog for activity recognition from video

CAVIAR-INRIA human activity dataset

28 videos
≈ 26.500 frames

- Separation between low-level events (LLE) and high-level events (HLE)
  - LLE: *walking, running, active, inactive, abrupt*
  - HLE: *meeting, moving, fighting, leaving_object*
- Probabilistic Logic approach: *Event Calculus in ProbLog* (Prob-EC) to infer the high-level events from an **algebra** of low-level events.
- Example:

$$\begin{gathered} initiatedAt(fighting(P_1, P_2) = true, T) \leftarrow \\ happensAt(abrupt(P_1), T), \\ holdsAt(close(P_1, P_2, 44) = true, T), \\ not\ happensAt(inactive(P_2), T). \end{gathered}$$

[Skarlatidis et al, TPLP 13]

Part V: Rule learning

Information Extraction in NELL

Recently-Learned Facts twitter

Refresh

instance	iteration	date learned	confidence
kelly_andrews is a female	826	29-mar-2014	98.7
investment_next_year is an economic sector	829	10-apr-2014	95.3
shibenik is a geopolitical entity that is an organization	829	10-apr-2014	97.2
quality_web_design_work is a character trait	826	29-mar-2014	91.0
mercedes_benz_cls_by_carlsson is an automobile manufacturer	829	10-apr-2014	95.2
social_work is an academic program at the university rutgers_university	827	02-apr-2014	93.8
dante wrote the book the_divine_comedy	826	29-mar-2014	93.8
willie_aames was born in the city los_angeles	831	16-apr-2014	100.0
kitt_peak is a mountain in the state or province arizona	831	16-apr-2014	96.9
greenwich is a park in the city london	831	16-apr-2014	100.0

instances for many different relations

degree of certainty

NELL: http://rtw.ml.cmu.edu/rtw/

Rule learning in NELL (1)

- Original approach
  - Make probabilistic data deterministic
  - run classic rule-learner (variant of FOIL)
  - re-introduce probabilities on learned rules and predict

Rule learning in NELL (2)

- Newer Page Rank Based Approach (Cohen et al.) -- ProPPR
  - Change the underlying model, from random graph / database to random walk one;
  - No longer "degree of belief" assigned to facts;
  - more like stochastic logic programs
  - Learn rules / parameters

Probabilistic Rule Learning

- Learn the rules directly in a PLP setting
- Generalize relational learning and inductive logic programming directly towards probabilistic setting
- Traditional rule learning/ILP as a special case
- Apply to probabilistic databases like NELL

Quinlan's Playtennis

ex	outlook ok	temperature ok	humidity ok	wind ok		class
1	t	t	f	f	-	+
2	f	t	f	t	-	+
3	t	f	f	f	-	-
4	f	f	t	f	-	-
...						
...						

Our Windsurfing Example

ex	pop	windok	sunshine	class
1	0,7	0,5	0,7	0,9
2	0,6	0,7	0,6	0,85
3	0,4	0,3	0,4	0,45
4	0,3	0,7	0,2	0,3
...				
...				

pop = Probability of Precipitation

Differences

- Observations (features) are uncertain
- Class is uncertain as well

- This type of data occurs naturally in applications in
  - image / video analysis
  - text processing
  - life sciences (e.g., Muggleton et al. MLJ 09)
  - probabilistic databases

Rule learning

In the logical setting

playtennis :- outlook=ok, wind=ok

playtennis :- outlook=ok, humidity=ok

In the probabilistic case

surfing :- not pop, windok

surfing :- not pop, sunshine

both a declarative and a probabilistic interpretation

In ProbLog (2)

Extended Setting

p1:: surfing(X) :- not pop(X) and windok(X).

p2:: surfing(X) :- not pop(X) and sunshine(X).

0.2::pop(e1). 0.7::windok(e1). 0.6::sunshine(e1).

H

B

e

?-P(surfing(e1)).

gives 0.8 x 0.7 x p1 + 0.8 x 0.6 x 0.3 x p2 = P(B U H |= e)

Inductive Probabilistic Logic Programs

Given

a set of example facts e ∈ E together with the probability p that they hold

a background theory B in ProbLog

a hypothesis space L (a set of clauses)

Find

$$\arg\min_H loss(H, B, E) = \arg\min_H \sum_{e_i \in E} |P_s(B \cup H \models e) - p_i|$$

Observations

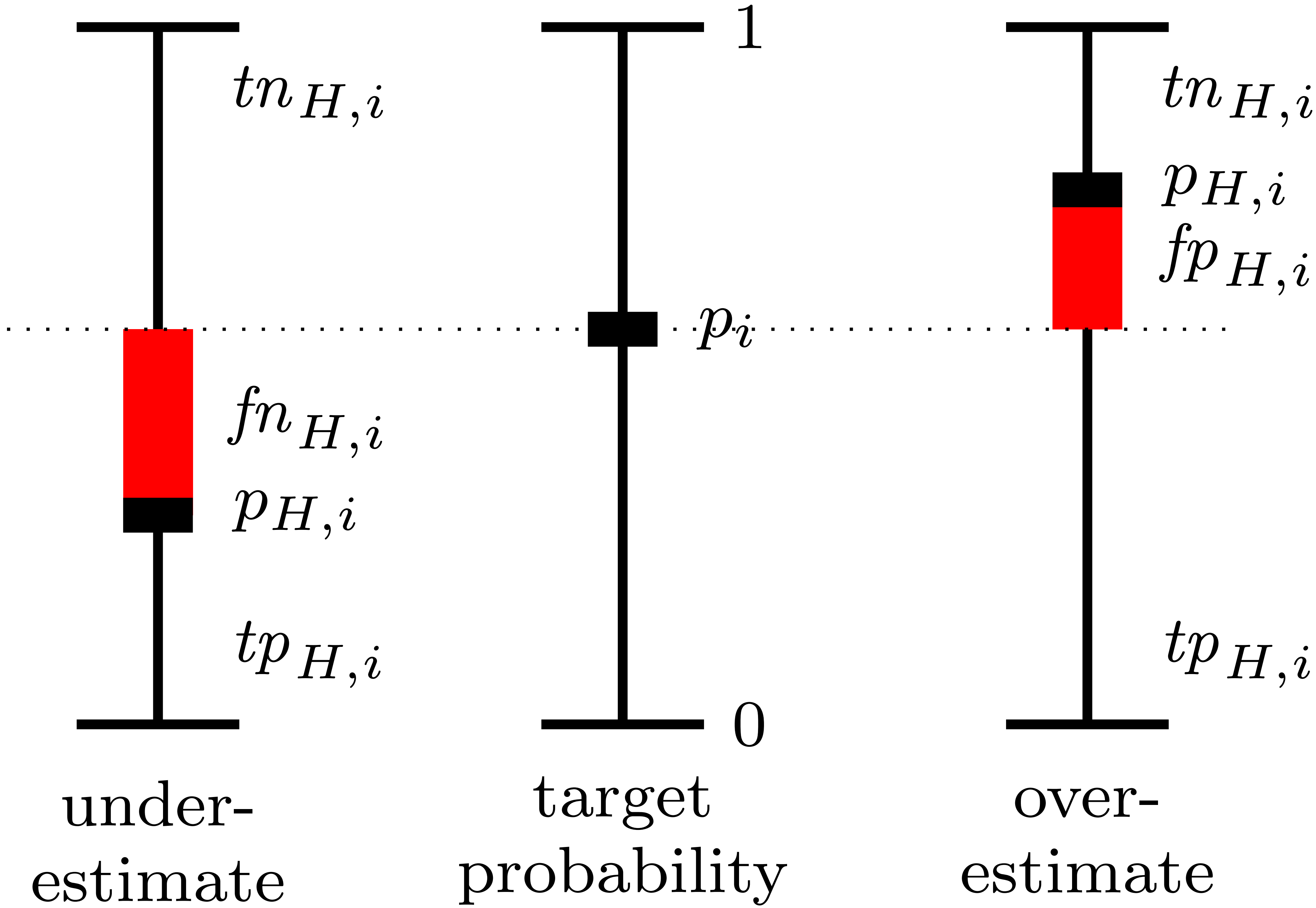
Propositional versus first order

- traditional rule learning = propositional
- inductive logic programming = first order

Deterministic case

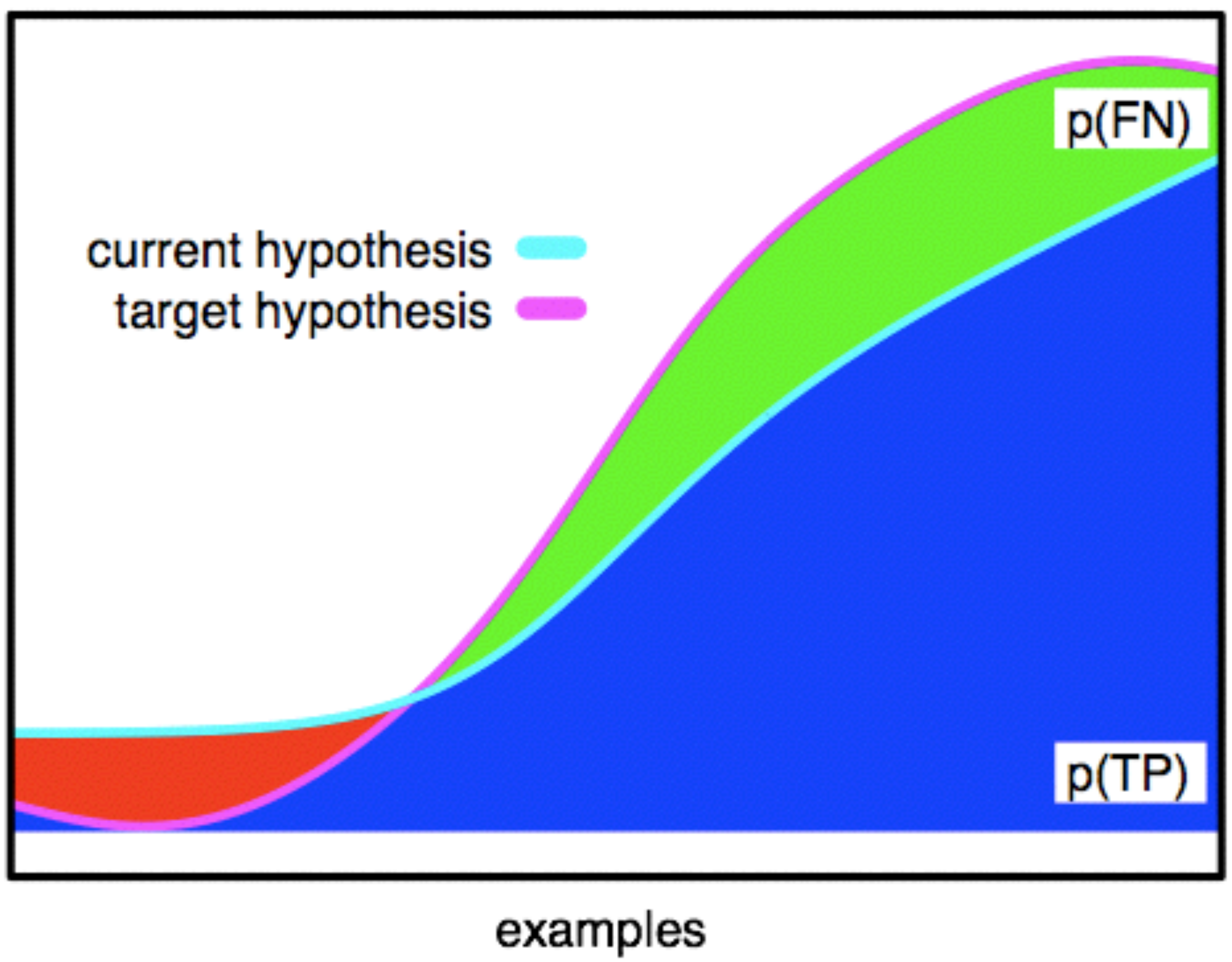
- all probabilities 0 or 1
- traditional rule learning / ILP as special case

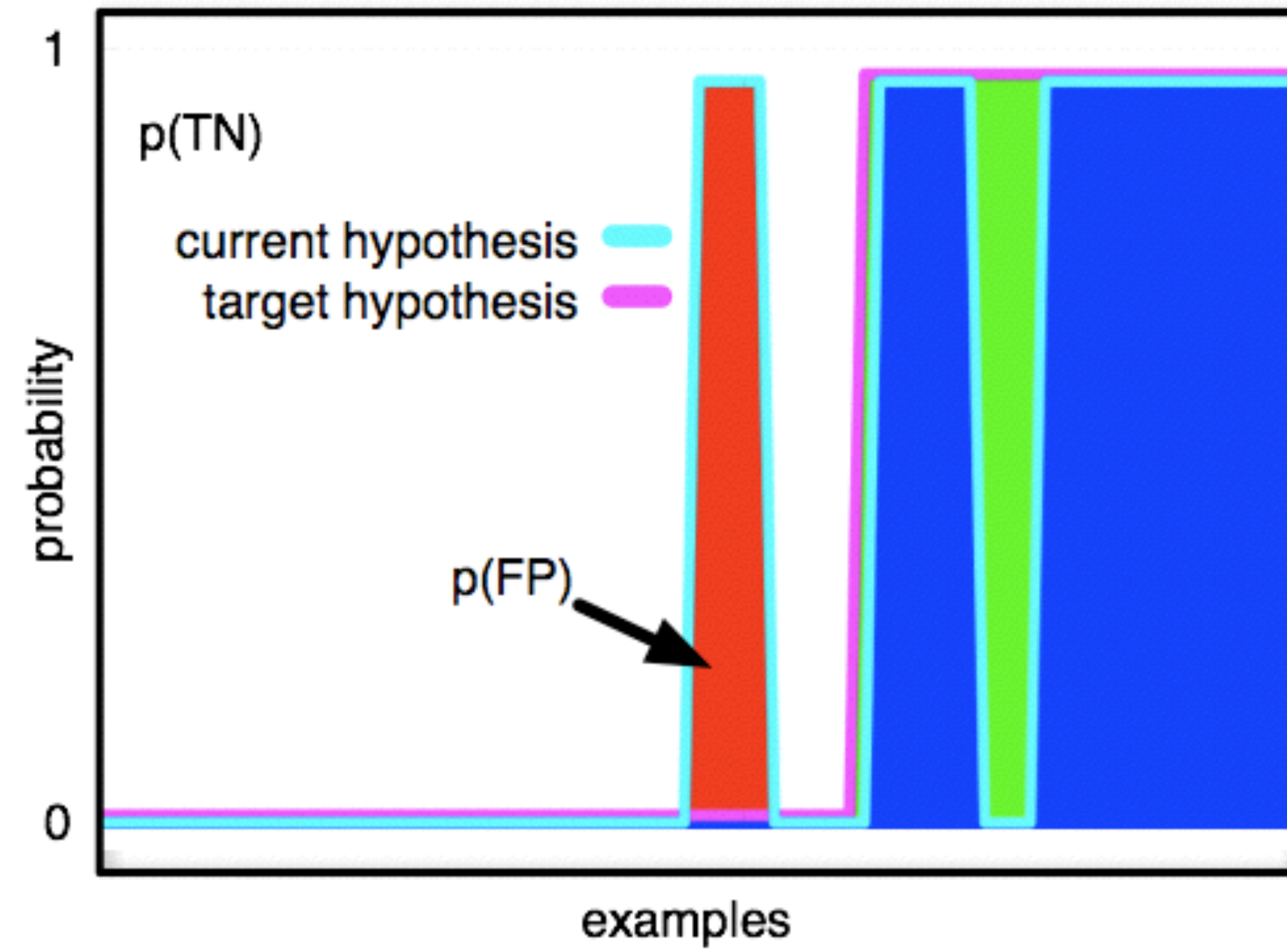
Analysis



1
$tn_{H,i}$
$p_{H,i}$
$fp_{H,i}$
p_i
$fn_{H,i}$
$p_{H,i}$
$tp_{H,i}$
$tp_{H,i}$
0
under-
estimate
target
probability
over-
estimate

Analysis


p(FN)
current hypothesis
target hypothesis
p(TP)
examples


1
p(TN)
current hypothesis
target hypothesis
probability
p(FP)
0
examples

Rule learning

Interesting properties

- adding a rule is monotonic, this can only increase the probability of an example
- adding a condition to a rule is anti-monotonic, this can only decrease the probability of an example
- several rules may be needed to cover an example
  - use all examples all of the time (do not delete them while learning), do not forget the positives
  - disjoint sum problem

ProbFOIL

Quinlan's well-known FOIL algorithm combined with ProbLog and probabilistic examples and background knowledge

Essentially a vanilla sequential covering algorithm with m-estimate as local score and accuracy as global score.

Criteria

$$
\begin{aligned}
precision &= \frac{TP}{TP + FP} \\
m\text{-}estimate &= \frac{TP + m \cdot \frac{P}{N}}{TP + FP + m} \\
recall &= \frac{TP}{TP + FN} \\
accuracy &= \frac{TP + TN}{TP + TN + FP + FN}
\end{aligned}
$$

local score

global score

Avoiding overfitting using significance test

ProbFOIL

Algorithm 1 The ProbFOIL^{+} learning algorithm

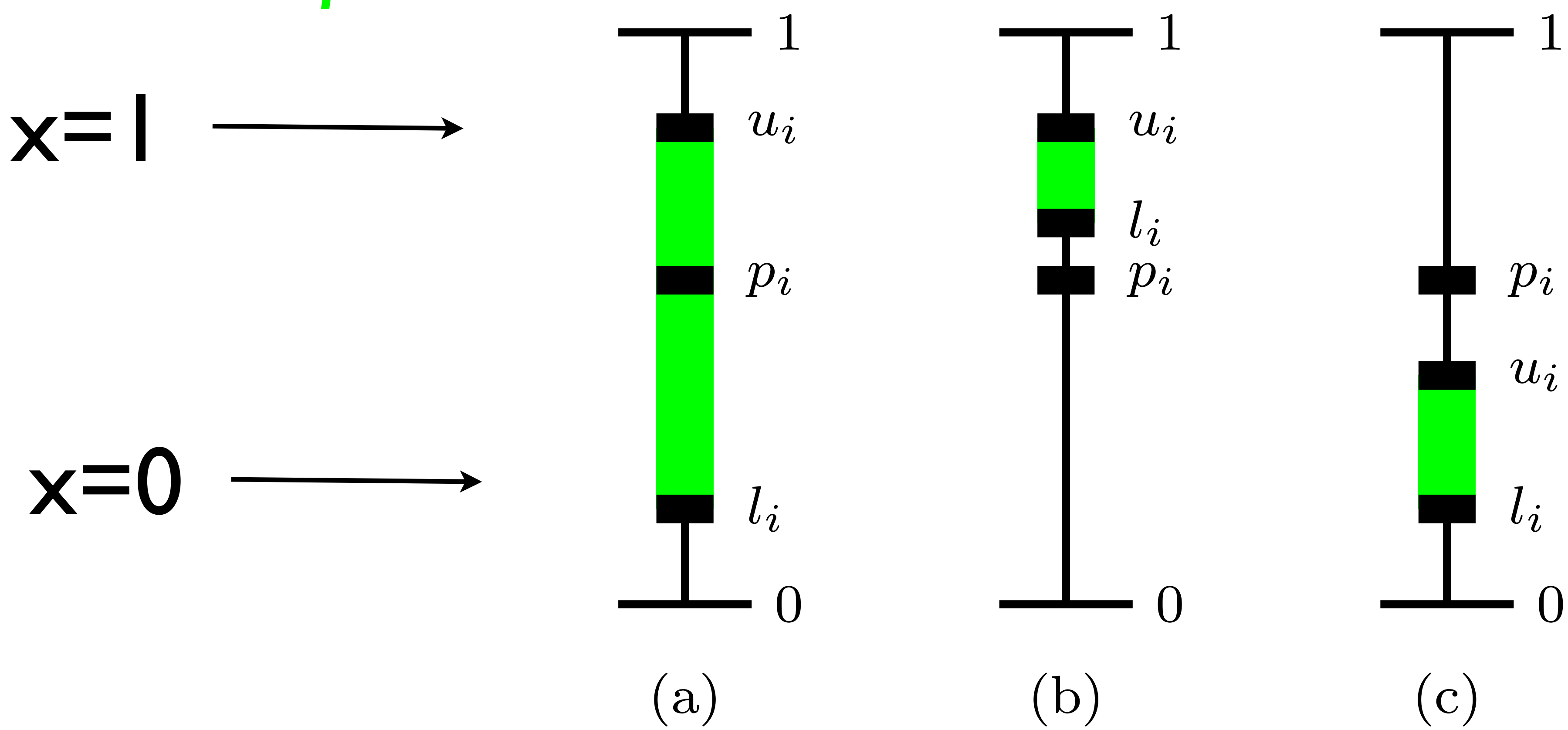
function ProbFOIL^{+}($target$) ▷ $target$ is the target predicate
  $H := \emptyset$
  while true **do**
    $clause :=$ LearnRule(H, $target$)
    if GlobalScore(H) $<$ GlobalScore($H \cup \{clause\}$) **then**
      $H := H \cup \{clause\}$
    else
      return H
function LearnRule(H, $target$)
  $candidates := \{x :: target \leftarrow true\}$ ▷ Start with an empty (probabilistic) body
  $bestrule := (x :: target \leftarrow true)$
  while $candidates \neq \emptyset$ **do** ▷ Grow rule
    $nextcandidates := \emptyset$
    for all $x :: target \leftarrow body \in candidates$ **do**
      for all $literal \in \rho(target \leftarrow body)$ **do** ▷ Generate all refinements
        if not RejectRefinement(H, $bestrule$, $x :: target \leftarrow body$) **then** ▷ Reject unsuited refinements
          $nextcandidates := nextcandidates \cup \{x :: target \leftarrow body \wedge l\}$
          if LocalScore (H, $x :: target \leftarrow body \wedge literal$) $>$ LocalScore(H, $bestrule$) **then**
            $bestrule := (x :: target \leftarrow body \wedge literal)$ ▷ Update best rule
    $candidates := nextcandidates$
  return $bestrule$

Extended rule learning

Learn rules with probability *x:: head :- body*

What changes ?

- value of *x* *determines prob. of coverage of example*



Extended rule learning

Express local score as a function of *x*

Compute optimal value of *x*

Implementation Optimizations

Incremental grounding

Simplified CNF conversion to ProbLog

Sometimes direct calculation of probabilities

Even simpler when propositional data only

Some language bias (range-restricted)

NELL

Table 5: Number of facts per predicate (NELL athlete dataset)

athletecoach(person,person)	18	athleteplaysforteam(person,team)	721
athleteplayssport(person,sport)	1921	teamplaysinleague(team,league)	1085
athleteplaysinleague(person,league)	872	athletealsoknownas(person,name)	17
coachesinleague(person,league)	93	coachesteam(person,team)	132
teamhomestadium(team,stadium)	198	teamplayssport(team,sport)	359
athleteplayssportsteamposition(person,position)	255	athletehomestadium(person,stadium)	187
athlete(person)	1909	attraction(stadium)	2
coach(person)	624	female(person)	2
male(person)	7	hobby(sport)	5
organization(league)	1	person(person)	2
personafrica(person)	1	personasia(person)	4
personaustralia(person)	22	personcanada(person)	1
personeurope(person)	1	personmexico(person)	108
personus(person)	6	sport(sport)	36
sportsleague(league)	18	sportsteam(team)	1330
sportsteamposition(position)	22	stadiumoreventvenue(stadium)	171

athleteplaysforteam

athleteplaysforteam(A,B) :- coachesteam(A,B).
0.875::athleteplaysforteam(A,B) :- teamhomestadium(B,C), athletehomestadium(A,C).
0.99080::athleteplaysforteam(A,B) :- teamhomestadium(B,_), male(A), athleteplayssport(A,_).
0.75::athleteplaysforteam(A,B) :- teamhomestadium(B,_), athleteplaysinleague(A,C), teamplaysinleague(B,C), athlete(A).
0.75::athleteplaysforteam(A,B) :- teamplayssport(B,C), athleteplayssport(A,C), coach(A), teamplaysinleague(B,_).
0.97555::athleteplaysforteam(A,B) :- personus(A), teamplayssport(B,_).
0.762::athleteplaysforteam(A,B) :- teamplayssport(B,C), athleteplayssport(A,C), personmexico(A), teamplaysinleague(B,_).
0.52571::athleteplaysforteam(A,B) :- teamplayssport(B,C), athleteplayssport(A,C), athleteplaysinleague(A,_), teamplaysinleague(B,_), athlete(A), teamplayssport(B,C).
0.50546::athleteplaysforteam(A,B) :- teamplayssport(B,_), teamplaysinleague(B,C), athleteplaysinleague(A,C), athleteplayssport(A,_).
0.50::athleteplaysforteam(A,B) :- teamplayssport(B,_), teamplaysinleague(B,C), athleteplaysinleague(A,C).
0.52941::athleteplaysforteam(A,B) :- teamplayssport(B,_), teamhomestadium(B,_), coach(A), teamplaysinleague(B,_).
0.55287::athleteplaysforteam(A,B) :- teamplayssport(B,_), teamplaysinleague(B,C), athleteplaysinleague(A,C), athlete(A).
0.46875::athleteplaysforteam(A,B) :- teamplayssport(B,_), teamplaysinleague(B,_), coach(A), teamhomestadium(B,_).

NELL

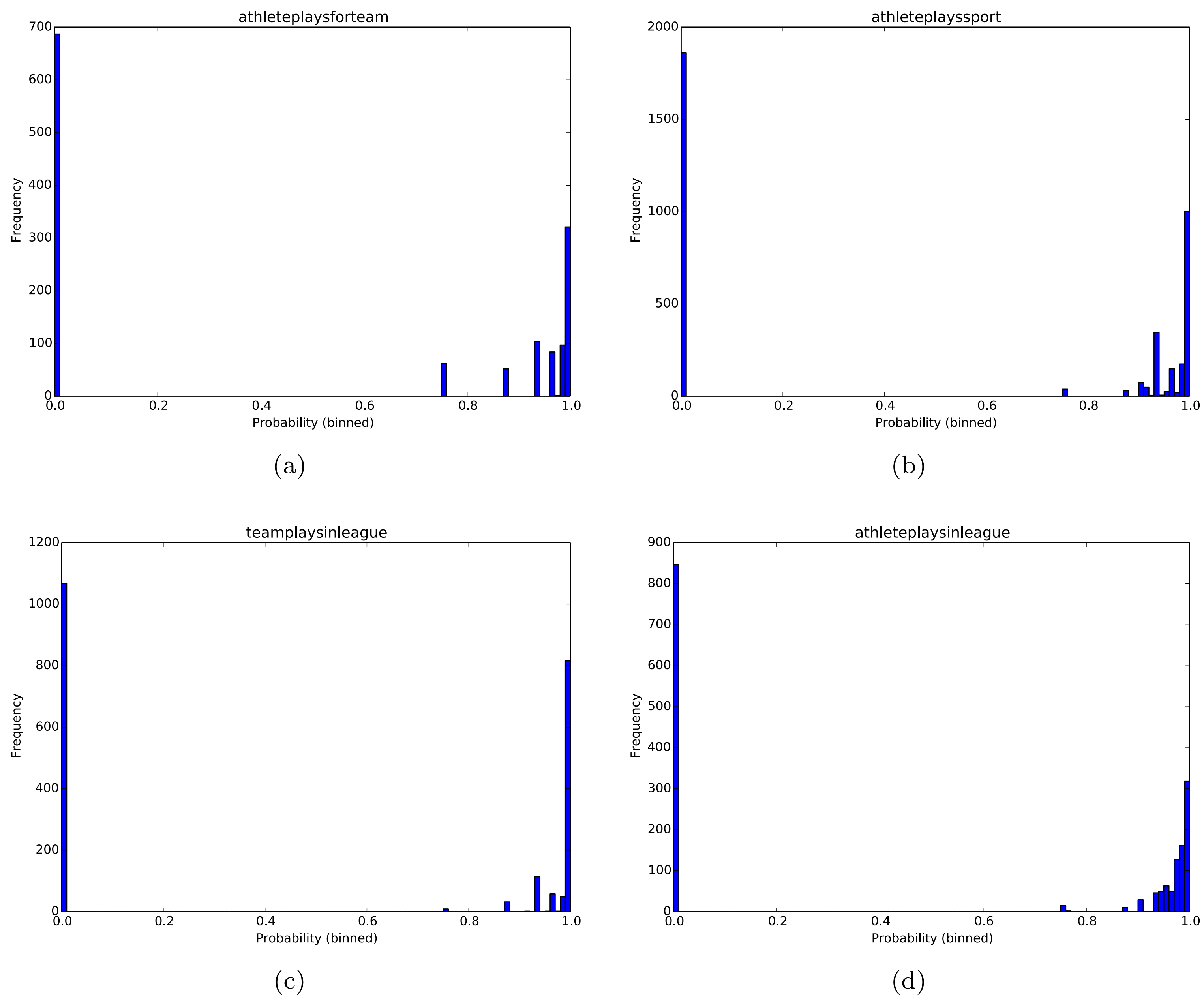


Fig. 5: Histogram of probabilities for each of the binary predicates with more then 500 facts: (a) athleteplaysforteam; (b) athleteplayssport; (c) teamplaysinleague; and, (d) athleteplaysinleague.

Contributions

Learning rules (or inducing logic programs) from uncertain/ probabilistic data

A new problem formulation

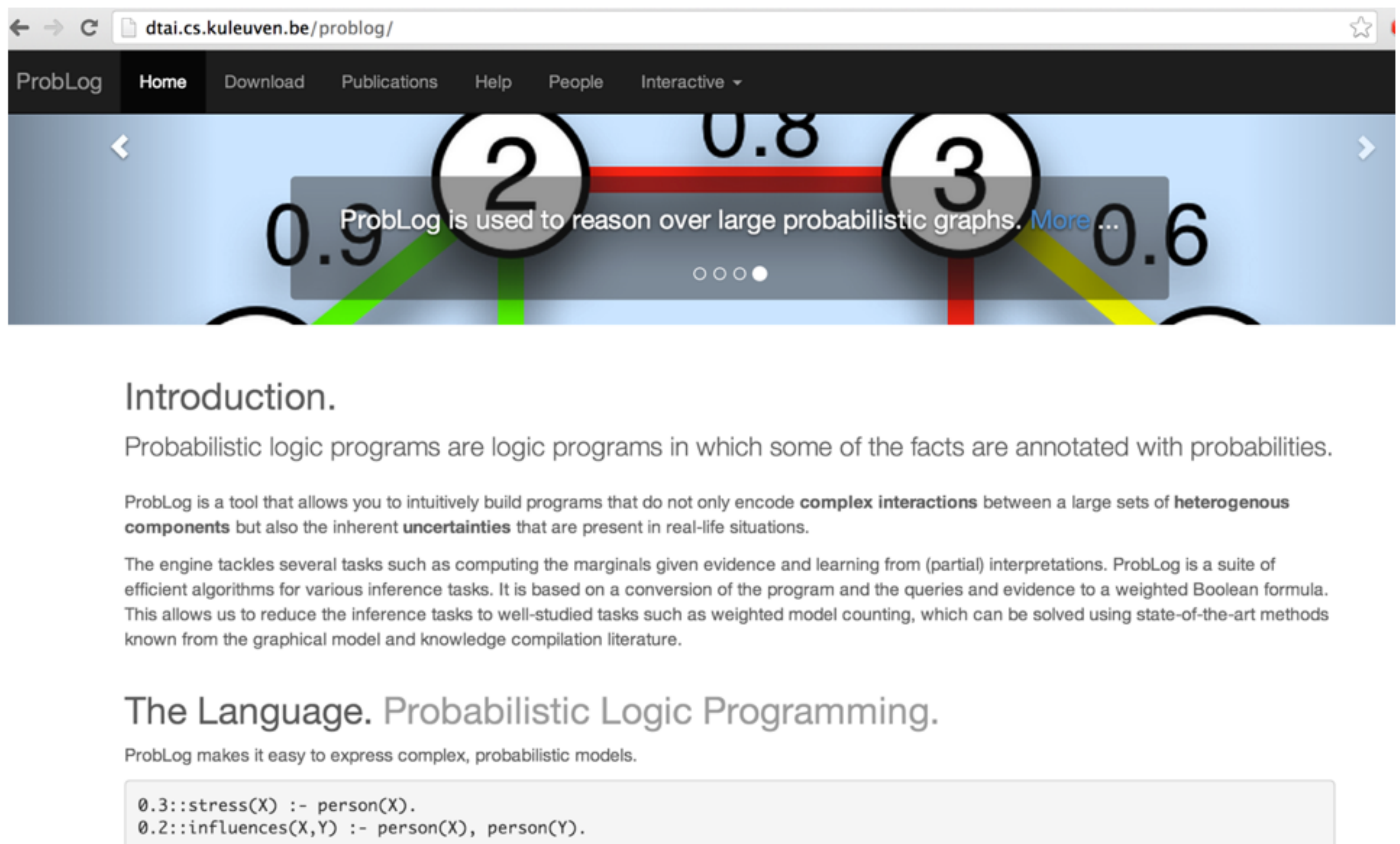
Traditional rule learning (ILP) is the deterministic special case

Traditional rule learning principles apply directly (including ROC analysis)

Maurice Bruynooghe
Bart Demoen
Anton Dries
Daan Fierens
Jason Filippou
Bernd Gutmann
Manfred Jaeger
Gerda Janssens
Kristian Kersting
Angelika Kimmig
Theofrastos Mantadelis
Wannes Meert
Bogdan Moldovan
Siegfried Nijssen
Davide Nitti
Joris Renkens
Kate Revoredo
Ricardo Rocha
Vitor Santos Costa
Dimitar Shterionov
Ingo Thon
Hannu Toivonen
Guy Van den Broeck
Mathias Verbeke
Jonas Vlasselaer

Thanks !

http://dtai.cs.kuleuven.be/problog



PLP Systems

- **PRISM** http://sato-www.cs.titech.ac.jp/prism/
- **ProbLog2** http://dtai.cs.kuleuven.be/problog/
- Yap Prolog http://www.dcc.fc.up.pt/~vsc/Yap/ includes
  - **ProbLog1**
  - **cplint** https://sites.google.com/a/unife.it/ml/cplint
  - **CLP(BN)**
  - **LP2**
- **PITA** in XSB Prolog http://xsb.sourceforge.net/
- **AILog2** http://artint.info/code/ailog/ailog2.html
- **SLPs** http://stoics.org.uk/~nicos/sware/pepl
- **contdist** http://www.cs.sunysb.edu/~cram/contdist/
- **DC** https://code.google.com/p/distributional-clauses
- **WFOMC** http://dtai.cs.kuleuven.be/ml/systems/wfomc

References

Bach SH, Broecheler M, Getoor L, O'Leary DP (2012) Scaling MPE inference for constrained continuous Markov random fields with consensus optimization. In: Proceedings of the 26th Annual Conference on Neural Information Processing Systems (NIPS-12)

Broecheler M, Mihalkova L, Getoor L (2010) Probabilistic similarity logic. In: Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence (UAI-10)

Bryant RE (1986) Graph-based algorithms for Boolean function manipulation. IEEE Transactions on Computers 35(8):677–691

Cohen SB, Simmons RJ, Smith NA (2008) Dynamic programming algorithms as products of weighted logic programs. In: Proceedings of the 24th International Conference on Logic Programming (ICLP-08)

Cussens J (2001) Parameter estimation in stochastic logic programs. Machine Learning 44(3):245–271

De Maeyer D, Renkens J, Cloots L, De Raedt L, Marchal K (2013) Phenetic: network-based interpretation of unstructured gene lists in e. coli. Molecular BioSystems 9(7):1594–1603

De Raedt L, Kimmig A (2013) Probabilistic programming concepts. CoRR abs/1312.4328

De Raedt L, Kimmig A, Toivonen H (2007) ProbLog: A probabilistic Prolog and its application in link discovery. In: Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI-07)

De Raedt L, Frasconi P, Kersting K, Muggleton S (eds) (2008) Probabilistic Inductive Logic Programming — Theory and Applications, Lecture Notes in Artificial Intelligence, vol 4911. Springer

Eisner J, Goldlust E, Smith N (2005) Compiling Comp Ling: Weighted dynamic programming and the Dyna language. In: Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP-05)

Fierens D, Blockeel H, Bruynooghe M, Ramon J (2005) Logical Bayesian networks and their relation to other probabilistic logical models. In: Proceedings of the 15th International Conference on Inductive Logic Programming (ILP-05)

Fierens D, Van den Broeck G, Bruynooghe M, De Raedt L (2012) Constraints for probabilistic logic programming. In: Proceedings of the NIPS Probabilistic Programming Workshop

Fierens D, Van den Broeck G, Renkens J, Shterionov D, Gutmann B, Thon I, Janssens G, De Raedt L (2014) Inference and learning in probabilistic logic programs using weighted Boolean formulas. Theory and Practice of Logic Programming (TPLP) FirstView

Getoor L, Friedman N, Koller D, Pfeffer A, Taskar B (2007) Probabilistic relational models. In: Getoor L, Taskar B (eds) An Introduction to Statistical Relational Learning, MIT Press, pp 129–174

Goodman N, Mansinghka VK, Roy DM, Bonawitz K, Tenenbaum JB (2008) Church: a language for generative models. In: Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence (UAI-08)

Gutmann B, Thon I, De Raedt L (2011a) Learning the parameters of probabilistic logic programs from interpretations. In: Proceedings of the 22nd European Conference on Machine Learning (ECML-11)

Gutmann B, Thon I, Kimmig A, Bruynooghe M, De Raedt L (2011b) The magic of logical inference in probabilistic programming. Theory and Practice of Logic Programming (TPLP) 11((4–5)):663–680

Huang B, Kimmig A, Getoor L, Golbeck J (2013) A flexible framework for probabilistic models of social trust. In: Proceedings of the International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction (SBP-13)

Jaeger M (2002) Relational Bayesian networks: A survey. Linköping Electronic Articles in Computer and Information Science 7(015)

Kersting K, Raedt LD (2001) Bayesian logic programs. CoRR cs.AI/0111058

Kimmig A, Van den Broeck G, De Raedt L (2011a) An algebraic Prolog for reasoning about possible worlds. In: Proceedings of the 25th AAAI Conference on Artificial Intelligence (AAAI-11)

Kimmig A, Demoen B, De Raedt L, Santos Costa V, Rocha R (2011b) On the implementation of the probabilistic logic programming language ProbLog. Theory and Practice of Logic Programming (TPLP) 11:235–262

Koller D, Pfeffer A (1998) Probabilistic frame-based systems. In: Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98)

McCallum A, Schultz K, Singh S (2009) FACTORIE: Probabilistic programming via imperatively defined factor graphs. In: Proceedings of the 23rd Annual Conference on Neural Information Processing Systems (NIPS-09)

Milch B, Marthi B, Russell SJ, Sontag D, Ong DL, Kolobov A (2005) Blog: Probabilistic models with unknown objects. In: Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI-05)

Moldovan B, De Raedt L (2014) Occluded object search by relational affordances. In: IEEE International Conference on Robotics and Automation (ICRA-14)

Moldovan B, Moreno P, van Otterlo M, Santos-Victor J, De Raedt L (2012) Learning relational affordance models for robots in multi-object manipulation tasks. In: IEEE International Conference on Robotics and Automation (ICRA-12)

Muggleton S (1996) Stochastic logic programs. In: De Raedt L (ed) Advances in Inductive Logic Programming, IOS Press, pp 254–264

Nitti D, De Laet T, De Raedt L (2013) A particle filter for hybrid relational domains. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-13)

Nitti D, De Laet T, De Raedt L (2014) Relational object tracking and learning. In: IEEE International Conference on Robotics and Automation (ICRA), June 2014

Pfeffer A (2001) IBAL: A probabilistic rational programming language. In: Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI-01)

Pfeffer A (2009) Figaro: An object-oriented probabilistic programming language. Tech. rep., Charles River Analytics

Poole D (2003) First-order probabilistic inference. In: Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI-03)

Richardson M, Domingos P (2006) Markov logic networks. Machine Learning 62(1-2):107–136

Santos Costa V, Page D, Cussens J (2008) CLP(BN): Constraint logic programming for probabilistic knowledge. In: De Raedt et al (2008), pp 156–188

Sato T (1995) A statistical learning method for logic programs with distribution semantics. In: Proceedings of the 12th International Conference on Logic Programming (ICLP-95)

Sato T, Kameya Y (2001) Parameter learning of logic programs for symbolic-statistical modeling. J Artif Intell Res (JAIR) 15:391–454

Sato T, Kameya Y (2008) New advances in logic-based probabilistic modeling by prism. In: Probabilistic Inductive Logic Programming, pp 118–155

Skarlatidis A, Artikis A, Filiopou J, Paliouras G (2014) A probabilistic logic programming event calculus. Theory and Practice of Logic Programming (TPLP) FirstView

Suciu D, Olteanu D, Ré C, Koch C (2011) Probabilistic Databases. Synthesis Lectures on Data Management, Morgan & Claypool Publishers

Taskar B, Abbeel P, Koller D (2002) Discriminative probabilistic models for relational data. In: Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence (UAI-02)

Thon I, Landwehr N, De Raedt L (2008) A simple model for sequences of relational state descriptions. In: Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases (ECML/PKDD-08)

Thon I, Landwehr N, De Raedt L (2011) Stochastic relational processes: Efficient inference and applications. Machine Learning 82(2):239–272

Van den Broeck G, Thon I, van Otterlo M, De Raedt L (2010) DTProbLog: A decision-theoretic probabilistic Prolog. In: Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI-10)

Van den Broeck G, Taghipour N, Meert W, Davis J, De Raedt L (2011) Lifted probabilistic inference by first-order knowledge compilation. In: Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI-11)

Vennekens J, Verbaeten S, Bruynooghe M (2004) Logic programs with annotated disjunctions. In: Proceedings of the 20th International Conference on Logic Programming (ICLP-04)

Vennekens J, Denecker M, Bruynooghe M (2009) CP-logic: A language of causal probabilistic events and its relation to logic programming. Theory and Practice of Logic Programming (TPLP) 9(3):245–308

Wang WY, Mazaitis K, Cohen WW (2013) Programming with personalized pagerank: a locally groundable first-order probabilistic logic. In: Proceedings of the 22nd ACM International Conference on Information and Knowledge Management (CIKM-13)